# 附录 A 矩阵代数

本附录简要地介绍学习金融计量学所需的矩阵知识。由于篇幅限制，我们未给出文中涉及的多数重要结论的证明过程，但是会引用可以提供证明过程的文献供读者参考。矩阵理论是金融计量学、最优化理论、机器学习以及深度学习等课程的最重要数学基础，著名的线性代数教材 Strang (2016) 在前言中写道：“The century of data has begun! ...... The truth is that vectors and matrices have become the language to know”，可见向量和矩阵知识的重要性。本附录的写作主要参考了 Calafiore et al. (2014)，此书的特点是突出强调向量和矩阵相关概念以及运算的几何意义，推荐读者进一步阅读此书以便加深理解。[1]

## A.1 向量

向量（vector）被定义为由有限个元素组成的列，例如，由 $a_1, \ldots, a_n$ 构成的 $n \times 1$ 维向量表示为：

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

$a_i$ 称为向量 $\boldsymbol{a}$ 的第 $i$ 个元素（或第 $i$ 个分量）。若 $\boldsymbol{a} \in \mathbb{R}^n$，则 $\boldsymbol{a}$ 可视为 $n$ 维欧氏空间中的一个点，此时，称 $a_i$ 为 $\boldsymbol{a}$ 在第 $i$ 个坐标轴上的坐标。

$\boldsymbol{a}$ 的转置定义为：

$$\boldsymbol{a}' \equiv (a_1, \ldots, a_n),$$

即，由 $a_1, \ldots, a_n$ 构成的 $1 \times n$ 维行向量。本书按照文献中的惯例，在引入向量时将其记为行向量的转置，如，$\boldsymbol{a} = (a_1, \ldots, a_n)'$。

此外，为了体现随机变量与其实现值的区别，本书将随机变量构成的向量用英文正体加粗记号表示，如，$\mathbf{x}$，其每个分量的一个实现值构成的向量则表示为 $\boldsymbol{x}$。当 $n = 1$ 时，称 $1 \times 1$ 维的向量为标量（scalar），本书在使用符号表示标量时不进行加粗处理，如，$x$（非随机变量）或 $\mathrm{x}$（随机变量）。

令 $\boldsymbol{a}^{(1)} = \left(a_1^{(1)}, \ldots, a_n^{(1)}\right)'$ 和 $\boldsymbol{a}^{(2)} = \left(a_1^{(2)}, \ldots, a_n^{(2)}\right)'$ 同为 $n \times 1$ 维的向量，分别定义 $\boldsymbol{a}^{(1)}$ 与 $\boldsymbol{a}^{(2)}$ 的和与差为：

$$\begin{aligned} \boldsymbol{a}^{(1)} + \boldsymbol{a}^{(2)} &= \left(a_1^{(1)} + a_1^{(2)}, \ldots, a_n^{(1)} + a_n^{(2)}\right)', \\ \boldsymbol{a}^{(1)} - \boldsymbol{a}^{(2)} &= \left(a_1^{(1)} - a_1^{(2)}, \ldots, a_n^{(1)} - a_n^{(2)}\right)'. \end{aligned}$$

令 $\alpha$ 为标量，$\alpha$ 与向量 $\boldsymbol{a} = (a_1, \ldots, a_n)'$ 的数乘定义为：

$$\alpha \boldsymbol{a} = (\alpha a_1, \ldots, \alpha a_n)'.$$

[1] 为了内容的相对完整，本附录会涉及本书正文未涉及的矩阵代数知识，读者可以有选择地阅读。

向量空间是由向量构成的集合，满足对向量加法和数乘运算的封闭性，关于向量空间的详细定义请参考 Horn et al. (2013) 的第 1 页[2]。向量空间的一个常见例子就是 $\mathbb{R}^n$。若向量空间 $\mathcal{X}$ 的非空子集 $\mathcal{V}$ 满足：对任意的标量 $\alpha$ 和 $\beta$，

$$\boldsymbol{x},\ \boldsymbol{y} \in \mathcal{V} \quad \Rightarrow \quad \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{V},$$

则称 $\mathcal{V}$ 为 $\mathcal{X}$ 的子空间。注意子空间总是包含 $\mathbf{0}$ 向量的。考虑一组向量构成的集合 $S = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$。将 $S$ 中的向量的所有线性组合构成的集合称为由 $S$ 张成的子空间，记为 $\operatorname{span}(S)$，即，

$$\operatorname{span}(S) \equiv \left\{\alpha_1\boldsymbol{x}^{(1)} + \cdots + \alpha_m\boldsymbol{x}^{(m)} \mid \alpha_1, \ldots, \alpha_m \text{为标量}\right\}.$$

在 $\mathbb{R}^n$ 中，当 $S = \{\boldsymbol{x}^{(1)}\}$ 时，$\operatorname{span}(S)$ 为沿着 $\boldsymbol{x}^{(1)}$ 的方向并且经过原点的一条直线，如图 A.1 所示；当 $S = \{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\}$ 且 $\boldsymbol{x}^{(1)}$ 与 $\boldsymbol{x}^{(2)}$ 不共线时，$\operatorname{span}(S)$ 为经过原点、$\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(2)}$ 的平面，如图 A.2 所示。[3]

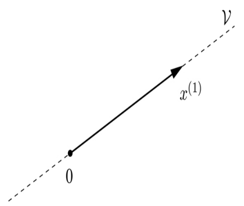


**图 A.1:** 由 $\boldsymbol{x}^{(1)}$ 张成的子空间

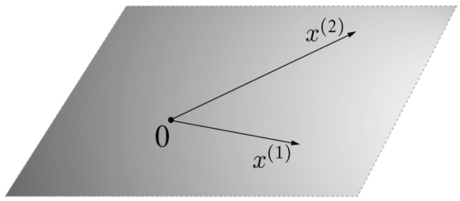


**图 A.2:** 由 $\boldsymbol{x}^{(1)}$ 和 $\boldsymbol{x}^{(2)}$ 张成的子空间

给定 $\mathbb{R}^n$ 中的两个子空间 $\mathcal{X}$ 和 $\mathcal{Y}$，$\mathcal{X}$ 与 $\mathcal{Y}$ 的直和，记为 $\mathcal{X} \oplus \mathcal{Y}$，定义为：

$$\{\boldsymbol{x} + \boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{y} \in \mathcal{Y}\}.$$

给定向量空间 $\mathcal{X}$ 中的一组向量 $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$，若满足

$$\sum_{i=1}^{m} \alpha_i \boldsymbol{x}^{(i)} = \mathbf{0} \quad \Rightarrow \quad \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)' = \mathbf{0},$$

则称 $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ 线性独立。对于前文定义的 $S$，若存在一组向量 $S^* = \{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d)}\}$，$d \leqslant m$，使得 $\operatorname{span}(S^*) = \operatorname{span}(S)$ 且 $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d)}$ 线性独立，则称 $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d)}$ 为 $\operatorname{span}(S)$ 的一组基，称 $d$ 为 $\operatorname{span}(S)$ 的维度。

与子空间密切相关的概念是仿射集，它本质上是子空间的平移。具体地，令 $\mathcal{V}$ 为 $\mathcal{X}$ 的子空间，由 $\mathcal{V}$ 生成的仿射集 $\mathcal{A}$ 定义为：

$$\mathcal{A} = \left\{\boldsymbol{x} \in \mathcal{X} \mid \boldsymbol{x} = \boldsymbol{v} + \boldsymbol{x}^{(0)},\ \boldsymbol{v} \in \mathcal{V}\right\},$$

其中，$\boldsymbol{x}^{(0)}$ 是给定的 $\mathcal{X}$ 中的点。直线是仿射集的一个例子，经过点 $\boldsymbol{x}_0$ 且与 $\boldsymbol{u}$ 平行的直线可以表示为：

$$\mathcal{L} = \{\boldsymbol{x} \in \mathcal{X} \mid \boldsymbol{x} = \boldsymbol{x}_0 + \lambda\boldsymbol{u},\ \lambda \in \mathbb{R}\}.$$

[2] 事实上，向量空间中的“向量”不仅仅局限于我们这里定义的向量。在现代数学中，任何满足对加法和数乘运算封闭且满足 Horn et al. (2013) 第 1 页列出的其他公理的数学对象都可以被当作是向量处理。例如，$n \times n$ 维实对称阵构成的集合是一个向量空间。

[3] 感谢 Calafiore et al. (2014) 的作者和剑桥出版社提供的图片。

## A.2 向量范数和内积

向量范数是欧氏空间中长度概念的推广，本质上向量范数是一个非负的实值函数，其定义如下：

**定义 A.1. 向量范数**

若从向量空间 $\mathcal{X}$ 到实数域 $\mathbb{R}$ 的一个非负映射 $\|\cdot\|$ 满足如下条件：

1. 正定性：对所有的 $\boldsymbol{x} \in \mathcal{X}$，有 $\|\boldsymbol{x}\| \geqslant 0$；并且，$\|\boldsymbol{x}\| = 0$ 当且仅当 $\boldsymbol{x} = \boldsymbol{0}$；
2. 齐次性：对所有的 $\boldsymbol{x} \in \mathcal{X}$ 和 $\alpha \in \mathbb{R}$，有 $\|\alpha\boldsymbol{x}\| = |\alpha| \cdot \|\boldsymbol{x}\|$；
3. 三角不等式：对所有的 $\boldsymbol{x}$，$\boldsymbol{y} \in \mathcal{X}$，有 $\|\boldsymbol{x}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$，

则称这个映射为向量范数。♣

最常用的向量范数为 $\mathcal{X} = \mathbb{R}^n$ 上的 $\ell_p$ 范数，其定义如下：

$$\|\boldsymbol{x}\|_p \equiv \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}, 1 \leqslant p < \infty.$$

当 $p=2$ 时，向量的 $\ell_2$ 范数是它的欧氏长度：

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{k=1}^{n} x_k^2};$$

当 $p=1$ 时，向量的 $\ell_1$ 范数是它的各个元素的绝对值之和：

$$\|\boldsymbol{x}\|_1 = \sum_{k=1}^{n} |x_k|;$$

当 $p \to \infty$ 时，向量的 $\ell_\infty$ 范数是它的绝对值最大的元素：

$$\|\boldsymbol{x}\|_\infty = \max_{k=1,\dots,n} |x_k|.$$

除了求和，两个向量之间的另一个基本运算是向量的内积（inner product）：

**定义 A.2. 向量内积**

向量空间 $\mathcal{X}$ 上的内积是一个实值函数，它将任意一对 $\mathcal{X}$ 中的元素 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 映射为一个标量，记为 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$。内积满足如下公理：对任意的 $\boldsymbol{x}$，$\boldsymbol{y}$，$\boldsymbol{z} \in \mathcal{X}$ 和标量 $\alpha$，

1. 非负性：$\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0$；
2. 非退化性：$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ 当且仅当 $\boldsymbol{x} = \boldsymbol{0}$；
3. 线性：$\langle \alpha\boldsymbol{x}, \boldsymbol{y} \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{y} \rangle$，$\langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$；
4. 对称性：$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$.

称定义了内积的向量空间为内积空间。♣

$\mathbb{R}^n$ 上的标准内积定义为：

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}'\boldsymbol{y} = \sum_{k=1}^{n} x_k y_k.$$

注意到，在一个内积空间中，$\sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ 满足范数的定义，通常将其记为 $\|\boldsymbol{x}\|$。在定义了标

准内积的内积空间 $\mathbb{R}^n$ 中，我们有：

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \|\boldsymbol{x}\|_2.$$

令 $\boldsymbol{x}$，$\boldsymbol{y} \in \mathbb{R}^n$ 为两个非 $\mathbf{0}$ 向量，它们之间夹角 $\theta$ 的余弦为：[4]

$$\cos\theta = \frac{\boldsymbol{x}'\boldsymbol{y}}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2}.$$

当 $\boldsymbol{x}'\boldsymbol{y} = 0$ 时，$\cos\theta = 0$，此时称这两个向量正交。因为 $|\cos\theta| \leqslant 1$，我们有：

**定理 A.1. 柯西-施瓦茨不等式**

对任意的 $\boldsymbol{x}$，$\boldsymbol{y} \in \mathbb{R}^n$，

$$|\boldsymbol{x}'\boldsymbol{y}| \leqslant \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2.$$

一般地，令 $\boldsymbol{x}$，$\boldsymbol{y} \in \mathcal{X}$，

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leqslant \|\boldsymbol{x}\| \|\boldsymbol{y}\|.$$

上面两个式子都在 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 共线时取等号。 ♡

对于 $n$ 维随机向量 $\mathbf{x}$ 和 $\mathbf{y}$，若将它们的内积定义为：

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbb{E}(|\mathbf{x}'\mathbf{y}|),$$

则有概率论版本的柯西-施瓦茨不等式：

$$\mathbb{E}(|\mathbf{x}'\mathbf{y}|) \leqslant \left(\mathbb{E}\|\mathbf{x}\|^2\right)^{1/2} \left(\mathbb{E}\|\mathbf{y}\|^2\right)^{1/2}. \tag{A.1}$$

## A.3 正交与正交补

前文提到的正交概念可以推广到一般的内积空间：如果内积空间 $\mathcal{X}$ 中的两个向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 满足 $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$，则称 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 正交，记为 $\boldsymbol{x} \perp \boldsymbol{y}$。考虑 $d$ 个非零向量 $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d)}$，如果对任意的 $1 \leqslant i \neq j \leqslant d$，有 $\boldsymbol{x}^{(i)} \perp \boldsymbol{x}^{(j)}$，则称 $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d)}$ 是相互正交的。可以证明（见 Calafiore et al. (2014) 的命题 2.1）：相互正交的向量是线性独立的。

令 $S$ 是由一组向量构成的集合，即，$S = \left\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(d)}\right\}$，若这组向量满足：对任意的 $1 \leqslant i$，$j \leqslant d$，

$$\left\langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \right\rangle = \begin{cases} 0 & \text{如果 } i \neq j, \\ 1 & \text{如果 } i = j, \end{cases}$$

则称 $S$ 为标准正交向量组。

令 $\mathcal{S} \subseteq \mathcal{X}$，若向量 $\boldsymbol{x} \in \mathcal{X}$ 满足：对任意的 $\boldsymbol{s} \in \mathcal{S}$，$\boldsymbol{x} \perp \boldsymbol{s}$，则称 $\boldsymbol{x}$ 与 $\mathcal{S}$ 正交。$\mathcal{X}$ 中所有与 $\mathcal{S}$ 正交的向量构成的集合称为 $\mathcal{S}$ 的正交补，记为 $\mathcal{S}^{\perp}$。容易验证 $\mathcal{S} \cap \mathcal{S}^{\perp} = \mathbf{0}$。如果对任意的 $\boldsymbol{x} \in \mathcal{X}$，存在唯一的方式将 $\boldsymbol{x}$ 表示为 $\boldsymbol{x} = \boldsymbol{a} + \boldsymbol{b}$，其中，$\boldsymbol{a} \in \mathcal{A}$，$\boldsymbol{b} \in \mathcal{B}$，$\mathcal{A}$ 和 $\mathcal{B}$ 为两个子空间，则称 $\mathcal{X}$ 为 $\mathcal{A}$ 和 $\mathcal{B}$ 的直和，记为 $\mathcal{X} = \mathcal{A} \oplus \mathcal{B}$。子空间和其正交补具有如下性质（证明见 Calafiore et al. (2014) 的定理 2.1）：

[4] 在直角坐标系下的证明过程请见 Calafiore et al. (2014) 的第 32 页。

**定理 A.2. 正交分解定理**

若 $\mathcal{S}$ 是内积空间 $\mathcal{X}$ 的一个子空间，则有：

$$\mathcal{X} = \mathcal{S} \oplus \mathcal{S}^{\perp}.$$

♡

## A.4 矩阵

矩阵是由数字构成的矩形阵列，例如，一个包含 $m$ 行和 $n$ 列的矩阵可以表示为：

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

若 $\boldsymbol{A}$ 中的元素都为实数，则记为 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$。矩阵 $\boldsymbol{A}$ 的第 $i$ 行和第 $j$ 列位置上的元素记为 $[\boldsymbol{A}]_{ij}$，并将其称为 $\boldsymbol{A}$ 的第 $(i,j)$ 位置上的元素。在上述例子中，$[\boldsymbol{A}]_{ij} = a_{ij}$。矩阵 $\boldsymbol{A}$ 的转置 $\boldsymbol{A}'$ 定义为：

$$[\boldsymbol{A}']_{ij} \equiv [\boldsymbol{A}]_{ji},$$

即，矩阵 $\boldsymbol{A}$ 的转置是将其行和列进行调换得到的矩阵。若 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，则 $\boldsymbol{A}' \in \mathbb{R}^{n\times m}$。

两个 $m \times n$ 维矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的和定义为 $m \times n$ 维矩阵，其第 $(i,j)$ 位置上的元素是：

$$[\boldsymbol{A} + \boldsymbol{B}]_{ij} = [\boldsymbol{A}]_{ij} + [\boldsymbol{B}]_{ij}.$$

矩阵 $\boldsymbol{A}$ 与标量 $\alpha$ 的数乘定义为 $m \times n$ 维矩阵，其第 $(i,j)$ 位置上的元素是：

$$[\alpha\boldsymbol{A}]_{ij} = \alpha \cdot [\boldsymbol{A}]_{ij}.$$

有了加法和数乘的定义，可以将 $\mathbb{R}^{m\times n}$ 视为向量空间。

令 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\boldsymbol{B} \in \mathbb{R}^{n\times q}$，矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的乘积 $\boldsymbol{AB}$ 为 $m \times q$ 维矩阵，其第 $(i,j)$ 位置上的元素是：

$$[\boldsymbol{AB}]_{ij} = \sum_{k=1}^{n} [\boldsymbol{A}]_{ik} \cdot [\boldsymbol{B}]_{kj}.$$

一般情况下，矩阵乘积不满足交换律，即，$\boldsymbol{AB} \neq \boldsymbol{BA}$。此外，两个矩阵乘积的转置满足：

$$(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'.$$

令 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，当 $m = n$ 时，矩阵 $\boldsymbol{A}$ 称为 $n$ 阶方阵。最常用的 $n$ 阶方阵为 $n$ 阶单位阵，记为 $\mathbf{I}_n$，其主对角线上的元素都为 1，其余位置上的元素都为 0。单位阵在矩阵运算中扮演着 1 在实数运算中的角色。例如，$\boldsymbol{A}\mathbf{I}_n = \boldsymbol{A}$。

将 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 视为由 $n$ 个列向量组成或由 $m$ 个行向量组成有助于理解特定的矩阵运算。具体地，若将 $\boldsymbol{A}$ 理解为由 $n$ 个列向量组成，则 $\boldsymbol{A}$ 可以表示为：

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix},$$

其中，$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$；若将 $\boldsymbol{A}$ 理解为由 $m$ 个行向量组成，则 $\boldsymbol{A}$ 可以表示为：

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\alpha}_1' \\ \boldsymbol{\alpha}_2' \\ \vdots \\ \boldsymbol{\alpha}_m' \end{bmatrix},$$

其中，$\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m \in \mathbb{R}^n$，$\boldsymbol{\alpha}_i'$ 表示 $\boldsymbol{A}$ 的第 $i$ 行，$1 \leqslant i \leqslant m$。利用上述两种表示矩阵的方式，我们可以通过以下三种不同的方式理解矩阵乘积：

- 将 $\boldsymbol{AB}$ 看作 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的各列相乘：

$$\boldsymbol{AB} = \boldsymbol{A} \begin{bmatrix} \boldsymbol{b}_1 & \ldots & \boldsymbol{b}_q \end{bmatrix} = \begin{bmatrix} \boldsymbol{Ab}_1 & \ldots & \boldsymbol{Ab}_q \end{bmatrix}.$$

- 将 $\boldsymbol{AB}$ 看作 $\boldsymbol{A}$ 的各行与 $\boldsymbol{B}$ 相乘：

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{\alpha}_1' \\ \boldsymbol{\alpha}_2' \\ \vdots \\ \boldsymbol{\alpha}_m' \end{bmatrix} \boldsymbol{B} = \begin{bmatrix} \boldsymbol{\alpha}_1'\boldsymbol{B} \\ \boldsymbol{\alpha}_2'\boldsymbol{B} \\ \vdots \\ \boldsymbol{\alpha}_m'\boldsymbol{B} \end{bmatrix}.$$

- 将 $\boldsymbol{AB}$ 看作秩为 1 的矩阵之和：

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_2' \\ \vdots \\ \boldsymbol{\beta}_n' \end{bmatrix} = \sum_{i=1}^{n} \boldsymbol{a}_i \boldsymbol{\beta}_i',$$

  其中，$\boldsymbol{\beta}_i'$ 是 $\boldsymbol{B}$ 的第 $i$ 行。

此外，矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 与向量 $\boldsymbol{b} \in \mathbb{R}^n$ 的乘积可以表示为：

$$\boldsymbol{Ab} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}' = \sum_{k=1}^{n} \boldsymbol{a}_k b_k.$$

上式表明，$\boldsymbol{Ab}$ 是 $\boldsymbol{A}$ 的列向量的线性组合，$\boldsymbol{b}$ 中的元素是此线性组合中 $\boldsymbol{A}$ 的各列的系数。同理，矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 左乘 $\boldsymbol{c} \in \mathbb{R}^m$ 的转置可以表示为：

$$\boldsymbol{c}'\boldsymbol{A} = \sum_{j=1}^{m} c_j \boldsymbol{\alpha}_j'.$$

上式表明，$\boldsymbol{c}'\boldsymbol{A}$ 是 $\boldsymbol{A}$ 的行向量的线性组合，$\boldsymbol{c}$ 中的元素是此线性组合中 $\boldsymbol{A}$ 的各行的系数。

对于 $n$ 阶方阵 $\boldsymbol{A}$，$\boldsymbol{A}$ 的迹，记为 $\operatorname{tr}\boldsymbol{A}$，定义为 $\boldsymbol{A}$ 的主对角线上的元素之和。令 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\boldsymbol{B} \in \mathbb{R}^{n\times m}$，迹的运算满足：

$$\operatorname{tr}\boldsymbol{AB} = \operatorname{tr}\boldsymbol{BA}. \tag{A.2}$$

更一般地，令 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\boldsymbol{B} \in \mathbb{R}^{n\times q}$，$\boldsymbol{C} \in \mathbb{R}^{q\times m}$，迹的运算满足：

$$\operatorname{tr}\boldsymbol{ABC} = \operatorname{tr}\boldsymbol{BCA} = \operatorname{tr}\boldsymbol{CAB}. \tag{A.3}$$

此外，迹还满足如下性质：

$$\begin{aligned} \operatorname{tr}(c\boldsymbol{A}) &= c\operatorname{tr}(\boldsymbol{A}), \\ \operatorname{tr}(\boldsymbol{A}') &= \operatorname{tr}(\boldsymbol{A}), \end{aligned}$$

$$\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) \quad = \quad \mathrm{tr}(\boldsymbol{A})+\mathrm{tr}(\boldsymbol{B}). \tag{A.4}$$

矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 的值域（range）定义为：

$$\mathcal{R}(\boldsymbol{A})=\{\boldsymbol{A}\boldsymbol{x}\mid\boldsymbol{x}\in\mathbb{R}^n\}.$$

根据此定义，矩阵的值域是其列向量张成的子空间。称值域 $\mathcal{R}(\boldsymbol{A})$ 的维度为 $\boldsymbol{A}$ 的秩，记为 $\mathrm{rank}(\boldsymbol{A})$。由子空间维度的定义，矩阵 $\boldsymbol{A}$ 的秩等于其线性独立的列的个数。可以证明（见 Abadir et al. (2005) 的 Exercise 4.5），矩阵 $\boldsymbol{A}$ 的秩也等于其线性独立的行的个数，即，$\mathrm{rank}(\boldsymbol{A})=\mathrm{rank}(\boldsymbol{A}')$。于是，$0\leqslant\mathrm{rank}(\boldsymbol{A})\leqslant\min(m,n)$。若 $\mathrm{rank}(\boldsymbol{A})=n$，称 $\boldsymbol{A}$ 为列满秩；若 $\mathrm{rank}(\boldsymbol{A})=m$，称 $\boldsymbol{A}$ 为行满秩。矩阵 $\boldsymbol{A}$ 的零空间（null space）定义为：

$$\mathcal{N}(\boldsymbol{A})=\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}\}.$$

线性代数基本定理建立了矩阵的零空间和其转置的值域之间的联系：（证明过程请参考 Calafiore et al. (2014) 的第 63 页）

**定理 A.3. 线性代数基本定理**

对任意给定的矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，$\mathcal{N}(\boldsymbol{A})^\perp=\mathcal{R}(\boldsymbol{A}')$，$\mathcal{R}(\boldsymbol{A})^\perp=\mathcal{N}(\boldsymbol{A}')$，因此，

$$\begin{aligned}\mathcal{N}(\boldsymbol{A})\oplus\mathcal{R}(\boldsymbol{A}') &= \mathbb{R}^n,\\ \mathcal{R}(\boldsymbol{A})\oplus\mathcal{N}(\boldsymbol{A}') &= \mathbb{R}^m,\end{aligned}$$

并且，

$$\begin{aligned}\dim\mathcal{N}(\boldsymbol{A})+\mathrm{rank}(\boldsymbol{A}) &= n,\\ \dim\mathcal{N}(\boldsymbol{A}')+\mathrm{rank}(\boldsymbol{A}) &= m,\end{aligned}$$

其中，$\dim\mathcal{S}$ 表示子空间 $\mathcal{S}$ 的维度。♡

## A.5 具有特殊结构的矩阵

### 对称矩阵

令 $\boldsymbol{A}$ 是 $n$ 阶方阵，若满足 $[\boldsymbol{A}]_{ij}=[\boldsymbol{A}]_{ji}$，$1\leqslant i,j\leqslant n$，则称 $\boldsymbol{A}$ 为 $n$ 阶对称矩阵。显然，对称矩阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}=\boldsymbol{A}'$。

### 对角矩阵

令 $\boldsymbol{A}$ 是 $n$ 阶方阵，若满足 $[\boldsymbol{A}]_{ij}=0$，$1\leqslant i\neq j\leqslant n$，则称 $\boldsymbol{A}$ 为 $n$ 阶对角矩阵。若 $\boldsymbol{A}$ 的主对角线上的元素构成的向量为 $\boldsymbol{a}=(a_{11},\ldots,a_{nn})'$，则 $\boldsymbol{A}$ 常表示为 $\mathrm{diag}\,(\boldsymbol{a})$。

### 三角矩阵

若 $n$ 阶方阵的主对角线以上元素都为 0，则称此类方阵为下三角矩阵；若 $n$ 阶方阵的主对角线以下元素都为 0，则称此类方阵为上三角矩阵。

### 正交矩阵

令 $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \end{bmatrix}$ 是 $n$ 阶方阵，若 $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ 构成 $\mathbb{R}^n$ 的一组标准正交基，即，

$$\boldsymbol{u}_i' \boldsymbol{u}_j = \begin{cases} 1 & \text{如果 } i = j, \\ 0 & \text{如果 } i \neq j, \end{cases}$$

则称 $\boldsymbol{U}$ 为正交矩阵。显然，正交矩阵 $\boldsymbol{U}$ 满足 $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}' = \mathbf{I}_n$。

向量左乘正交矩阵可以保持范数不变，事实上，对任意的 $\boldsymbol{x} \in \mathbb{R}^n$，

$$\|\boldsymbol{U}\boldsymbol{x}\|_2^2 = (\boldsymbol{U}\boldsymbol{x})'(\boldsymbol{U}\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{x} = \|\boldsymbol{x}\|_2^2.$$

此外，左乘正交矩阵还可以保持两个向量之间的夹角不变。令 $\boldsymbol{x}$，$\boldsymbol{y} \in \mathbb{R}^n$，

$$(\boldsymbol{U}\boldsymbol{x})'(\boldsymbol{U}\boldsymbol{y}) = \boldsymbol{x}'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{y} = \boldsymbol{x}'\boldsymbol{y}.$$

于是，$\boldsymbol{U}\boldsymbol{x}$ 和 $\boldsymbol{U}\boldsymbol{y}$ 之间的夹角 $\theta^*$ 的余弦为：

$$\cos\theta^* = \frac{\boldsymbol{U}\boldsymbol{x}'\boldsymbol{U}\boldsymbol{y}}{\|\boldsymbol{U}\boldsymbol{x}\|_2 \|\boldsymbol{U}\boldsymbol{y}\|_2} = \cos\theta.$$

### 幂等矩阵

如果 $n$ 阶方阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^2 = \boldsymbol{A}$，则称 $\boldsymbol{A}$ 为幂等矩阵。例如，本书正文第二章提到的投影矩阵 $\mathbf{P}_{\mathbf{X}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ 是幂等矩阵。同时，$\mathbf{P}_{\mathbf{X}}$ 也是对称矩阵。但是，并非所有的幂等矩阵都是对称矩阵，例如：考虑

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},$$

容易验证 $\boldsymbol{A}$ 是非对称的幂等矩阵。

### 并矢矩阵（向量的直积）

若 $\mathbb{R}^{m \times n}$ 中的矩阵 $\boldsymbol{A}$ 可以表示为 $\boldsymbol{A} = \boldsymbol{u}\boldsymbol{v}'$，其中，$\boldsymbol{u} \in \mathbb{R}^m$，$\boldsymbol{v} \in \mathbb{R}^n$，则称 $\boldsymbol{A}$ 为并矢矩阵或向量的直积。注意到 $[\boldsymbol{A}]_{ij} = u_i v_j$，因此，并矢矩阵的每一行是其他行的缩放，缩放比例由 $\boldsymbol{u}$ 确定；它的每一列是其他列的缩放，缩放比例由 $\boldsymbol{v}$ 确定。事实上（以行为例），$\boldsymbol{A}$ 的第 $i$ 和第 $j$ 行分别为 $u_i \cdot (v_1, \dots, v_n)$ 和 $u_j \cdot (v_1, \dots, v_n)$，于是，$\boldsymbol{A}$ 的第 $j$ 行可以表示为它的第 $i$ 行乘以 $u_j / u_i$。

此外，并矢矩阵 $\boldsymbol{A}$ 的值域为一条直线，这是因为：对于任意输入向量 $\boldsymbol{x} \in \mathbb{R}^n$，

$$\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{u}\boldsymbol{v}')\boldsymbol{x} = (\boldsymbol{v}'\boldsymbol{x})\boldsymbol{u}.$$

从上式可以看出，$\boldsymbol{A}$ 的值域是由 $\boldsymbol{u}$ 张成的直线。因为直线的维度是 1，所以 $\operatorname{rank}(\boldsymbol{A}) = 1$。

## A.6 矩阵范数

与向量范数类似，矩阵范数的定义如下：

**定义 A.3.（广义）矩阵范数**

若从 $\mathbb{R}^{m\times n}$ 到实数域 $\mathbb{R}$ 的一个非负映射 $\|\cdot\|$ 满足如下条件：

1. 正定性：对所有的 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，有 $\|\boldsymbol{A}\|\geqslant 0$；并且，$\|\boldsymbol{A}\|=0$ 当且仅当 $\boldsymbol{A}=\boldsymbol{0}$；
2. 齐次性：对所有的 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 和 $\alpha\in\mathbb{R}$，有 $\|\alpha\boldsymbol{A}\|=|\alpha|\cdot\|\boldsymbol{A}\|$；
3. 三角不等式：对所有的 $\boldsymbol{A}$, $\boldsymbol{B}\in\mathbb{R}^{m\times n}$，有 $\|\boldsymbol{A}+\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{B}\|$，

则称这个映射为广义矩阵范数。

若 $m=n$，并且广义矩阵范数 $\|\cdot\|$ 还满足如下性质：

4. 次可乘性：$\|\boldsymbol{AB}\|\leqslant\|\boldsymbol{A}\|\cdot\|\boldsymbol{B}\|$，

则称 $\|\cdot\|$ 为矩阵范数。♣

向量的 $\ell_p$ 范数可以容易地推广到矩阵的 $\ell_p$ 范数，常用的是 $p=1,2$ 的情形：

- 矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 的 $\ell_1$ 范数定义为：
$$\|\boldsymbol{A}\|_1=\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|,$$
即，$\|\boldsymbol{A}\|_1$ 为 $\boldsymbol{A}$ 中所有元素的绝对值之和。
- 矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 的 $\ell_2$ 范数称为 Frobenius 范数，其定义为：
$$\|\boldsymbol{A}\|_F=\sqrt{\operatorname{tr}(\boldsymbol{AA}')}=\sqrt{\sum_{i,j}a_{ij}^2}, \tag{A.5}$$
即，$\|\boldsymbol{A}\|_F$ 为 $\boldsymbol{A}$ 中所有元素的平方和开根号。显然，矩阵的 Frobenius 范数是向量 $\ell_2$ 范数的一个自然推广。本书提到矩阵范数时如无特殊说明都是指矩阵的 Frobenius 范数，并简单地记为 $\|\boldsymbol{A}\|$。

矩阵的 Frobenius 范数具有正交不变性，即，对任意的正交矩阵 $\boldsymbol{U}\in\mathbb{R}^{m\times m}$ 和 $\boldsymbol{V}\in\mathbb{R}^{n\times n}$，
$$\|\boldsymbol{UAV}\|_F^2=\operatorname{tr}(\boldsymbol{UAVV}'\boldsymbol{A}'\boldsymbol{U}')=\operatorname{tr}(\boldsymbol{AA}')=\|\boldsymbol{A}\|_F^2.$$
令 $\boldsymbol{A}$ 的行为 $\boldsymbol{\alpha}_1',\ldots,\boldsymbol{\alpha}_m'$，则有：
$$\|\boldsymbol{A}\|_F^2=\sum_{i=1}^{m}\|\boldsymbol{\alpha}_i'\|_2^2.$$
因此，对任意的 $\boldsymbol{x}\in\mathbb{R}^n$，有：
$$\|\boldsymbol{Ax}\|_2^2=\sum_{i=1}^{m}|\boldsymbol{\alpha}_i'\boldsymbol{x}|^2\leqslant\sum_{i=1}^{m}\|\boldsymbol{\alpha}_i'\|_2^2\|\boldsymbol{x}\|_2^2=\|\boldsymbol{A}\|_F^2\|\boldsymbol{x}\|_2^2,$$
其中，不等号成立用到了柯西-施瓦茨不等式。上式意味着：
$$\|\boldsymbol{Ax}\|_2\leqslant\|\boldsymbol{A}\|_F\|\boldsymbol{x}\|_2. \tag{A.6}$$
式 (A.6) 称为 Frobenius 范数与向量范数是相容的，利用这一结果可以验证矩阵的 Frobenius 范数满足次可乘性：令 $\boldsymbol{B}\in\mathbb{R}^{n\times q}$，并将 $\boldsymbol{B}$ 记为 $\begin{bmatrix}\boldsymbol{b}_1 & \ldots & \boldsymbol{b}_q\end{bmatrix}$，则有：
$$\|\boldsymbol{AB}\|_F^2=\sum_{j=1}^{q}\|\boldsymbol{Ab}_j\|_2^2\leqslant\sum_{j=1}^{q}\|\boldsymbol{A}\|_F^2\|\boldsymbol{b}_j\|_2^2=\|\boldsymbol{A}\|_F^2\|\boldsymbol{B}\|_F^2.$$
除了上述矩阵的 $\ell_p$ 范数之外，还可以定义矩阵的算子范数，算子范数是由向量范数诱导出的矩阵范数。

**定义 A.4. 算子范数**

设 $\|\cdot\|_\delta$ 和 $\|\cdot\|_\gamma$ 分别是向量空间 $\mathbb{R}^n$ 和 $\mathbb{R}^m$ 上的向量范数。由下式定义的非负实值函数 $\|\cdot\|_{\gamma\delta}:\mathbb{R}^{m\times n}\to\mathbb{R}$:

$$\|\boldsymbol{A}\|_{\gamma\delta}=\max_{\boldsymbol{u}\neq\boldsymbol{0}}\frac{\|\boldsymbol{A}\boldsymbol{u}\|_\gamma}{\|\boldsymbol{u}\|_\delta}=\max_{\|\boldsymbol{u}\|_\delta=1}\|\boldsymbol{A}\boldsymbol{u}\|_\gamma$$

称为由向量范数 $\|\cdot\|_\delta$ 和 $\|\cdot\|_\gamma$ 诱导的广义矩阵范数，也称为算子范数。若 $\|\cdot\|_\delta$ 和 $\|\cdot\|_\gamma$ 都取为相应向量空间上的 $\ell_p$ 范数，则称算子范数为矩阵的 $p$ 范数。

根据算子范数的定义，它满足如下性质：

$$\|\boldsymbol{A}\boldsymbol{u}\|_\gamma\leqslant\|\boldsymbol{A}\|_{\gamma\delta}\|\boldsymbol{u}\|_\delta,$$

此性质称为 $\|\cdot\|_{\gamma\delta}$ 与 $\|\cdot\|_\delta$ 和 $\|\cdot\|_\gamma$ 是相容的。

常用的矩阵 $p$ 范数有：

- 1 范数：由向量的 $\ell_1$ 范数诱导的算子范数：

$$\|\boldsymbol{A}\|_1=\max_{\|\boldsymbol{u}\|_1=1}\|\boldsymbol{A}\boldsymbol{u}\|_1.$$

- 2 范数：由向量的 $\ell_2$ 范数诱导的算子范数：

$$\|\boldsymbol{A}\|_2=\max_{\|\boldsymbol{u}\|_2=1}\|\boldsymbol{A}\boldsymbol{u}\|_2.$$

- $\infty$ 范数：由向量的 $\ell_\infty$ 范数诱导的算子范数：

$$\|\boldsymbol{A}\|_\infty=\max_{\|\boldsymbol{u}\|_\infty=1}\|\boldsymbol{A}\boldsymbol{u}\|_\infty.$$

可以证明：（见陈公宁 (2007) 的定理 4）

$$\begin{aligned}\|\boldsymbol{A}\|_1&=\max_{j=1,\dots,n}\sum_{i=1}^m|a_{ij}|,\quad\|\boldsymbol{A}\|_\infty=\max_{i=1,\dots,m}\sum_{j=1}^n|a_{ij}|,\\ \|\boldsymbol{A}\|_2&=\sqrt{\lambda_{\max}\left(\boldsymbol{A}'\boldsymbol{A}\right)},\end{aligned}$$

其中，$\lambda_{\max}(\cdot)$ 表示矩阵的最大特征值[5]，注意到矩阵 $\boldsymbol{A}'\boldsymbol{A}$ 的特征值是 $\boldsymbol{A}$ 的奇异值的平方[6]，因此，矩阵的 2 范数是该矩阵的最大奇异值。矩阵的 2 范数也称为矩阵的谱（spectral）范数。

与矩阵的 Frobenius 范数密切相关的概念是矩阵的 Frobenius 内积：令 $\boldsymbol{A}$, $\boldsymbol{B}\in\mathbb{R}^{m\times n}$, $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的 Frobenius 内积定义为：

$$\langle\boldsymbol{A},\boldsymbol{B}\rangle\equiv\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{B}'\right)=\sum_{i=1}^m\sum_{j=1}^n a_{ij}b_{ij}.\tag{A.7}$$

当 $\boldsymbol{A}=\boldsymbol{B}$ 时，$\langle\boldsymbol{A},\boldsymbol{B}\rangle$ 为 $\boldsymbol{A}$ 的 Frobenius 范数的平方。矩阵的内积也有对应的柯西-施瓦茨不等式：令 $\boldsymbol{A}$, $\boldsymbol{B}\in\mathbb{R}^{m\times n}$

$$|\langle\boldsymbol{A},\boldsymbol{B}\rangle|\leqslant\|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F,$$

当且仅当 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 线性相关时，上式可以取到等号。

[5] 见下一小节中关于特征值的定义。

[6] 见 A.13 中关于奇异值的定义。

## A.7 行列式、特征值和特征向量

任意矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 都可以看作是从 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的线性映射，这是因为对任意的 $\boldsymbol{x} \in \mathbb{R}^n$，$\boldsymbol{x} \longmapsto \boxed{\boldsymbol{A}} \longmapsto \boldsymbol{Ax} \in \mathbb{R}^m$；并且，

$$\boldsymbol{A}(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha\boldsymbol{Ax} + \beta\boldsymbol{Ay},$$

其中，$\alpha$, $\beta$ 是标量，$\boldsymbol{y} \in \mathbb{R}^n$，即，$\boldsymbol{A}(\boldsymbol{x}) \equiv \boldsymbol{Ax}$ 是线性映射。若 $\boldsymbol{A}$ 为 $n$ 阶方阵，则它可以看作是 $\mathbb{R}^n$ 到其自身的线性映射。

为了介绍特征值等概念及其几何意义，本节先考虑 $\mathbb{R}^n$ 中过原点的直线经过 $n$ 阶方阵 $\boldsymbol{A}$ 映射之后的效果。具体地，令 $\boldsymbol{u}$ 是 $\mathbb{R}^n$ 中的非零向量，过原点和 $\boldsymbol{u}$ 的直线记为：

$$\mathcal{L}_{\boldsymbol{u}} = \{\boldsymbol{x} = \alpha\boldsymbol{u} \mid \alpha \in \mathbb{R}\}.$$

对直线 $\mathcal{L}_{\boldsymbol{u}}$ 上的任意一点 $\boldsymbol{x}$，经过 $\boldsymbol{A}$ 映射之后得到：

$$\boldsymbol{y} = \boldsymbol{Ax} = \alpha\boldsymbol{Au}.$$

关于 $\boldsymbol{y}$ 的长度，我们有如下结论：

$$\|\boldsymbol{y}\|_2 = \|\boldsymbol{Ax}\|_2 = |\alpha|\|\boldsymbol{Au}\|_2 = \frac{\|\boldsymbol{Au}\|_2}{\|\boldsymbol{u}\|_2}|\alpha|\|\boldsymbol{u}\|_2 = \gamma_{\boldsymbol{u}}\|\boldsymbol{x}\|_2,$$

其中，$\gamma_{\boldsymbol{u}} = \|\boldsymbol{Au}\|_2/\|\boldsymbol{u}\|_2$。类似地，$\boldsymbol{y}$ 与 $\boldsymbol{x}$ 之间夹角 $\theta_{\boldsymbol{u}}$ 的余弦为：

$$\cos\theta_{\boldsymbol{u}} = \frac{\boldsymbol{y}'\boldsymbol{x}}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2} = \frac{\boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x}}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2} = \frac{\boldsymbol{u}'\boldsymbol{A}'\boldsymbol{u}}{\gamma_{\boldsymbol{u}}\|\boldsymbol{u}\|_2^2}.$$

注意到 $\gamma_{\boldsymbol{u}}$ 和 $\cos\theta_{\boldsymbol{u}}$ 都只与 $\boldsymbol{u}$ 和 $\boldsymbol{A}$ 有关，而与 $\boldsymbol{x}$ 在直线上的具体位置无关。因此，$\boldsymbol{Ax}$ 的作用是先将 $\boldsymbol{x}$ 旋转固定的角度 $\theta_{\boldsymbol{u}}$，再将旋转后的向量的长度缩放固定比例 $\gamma_{\boldsymbol{u}}$。此外，容易验证：在 $\boldsymbol{u}$ 上乘以任意非零的常数都不会改变 $\gamma_{\boldsymbol{u}}$ 和 $\cos\theta_{\boldsymbol{u}}$ 的取值，因此可以将 $\boldsymbol{u}$ 标准化为单位长度的向量。

### 矩阵的行列式

首先考虑 2 阶方阵：

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

其行列式定义为：

$$\det\boldsymbol{A} \equiv a_{11}a_{22} - a_{12}a_{21}.$$

为了说明上述定义的几何意义，取 $\mathbb{R}^2$ 中的单位正方形，其四个顶点分别为：$\boldsymbol{x}^{(1)} = (0,0)'$、$\boldsymbol{x}^{(2)} = (1,0)'$、$\boldsymbol{x}^{(3)} = (0,1)'$ 和 $\boldsymbol{x}^{(4)} = (1,1)'$。将这四个向量经过 $\boldsymbol{A}$ 进行映射之后得到：

$$\begin{aligned} \boldsymbol{y}^{(1)} &= \boldsymbol{Ax}^{(1)} = (0,0)'; \quad \boldsymbol{y}^{(2)} = \boldsymbol{Ax}^{(2)} = (a_{11}, a_{21})'; \\ \boldsymbol{y}^{(3)} &= \boldsymbol{Ax}^{(3)} = (a_{12}, a_{22})'; \quad \boldsymbol{y}^{(4)} = \boldsymbol{Ax}^{(4)} = (a_{11}+a_{12}, a_{21}+a_{22})'. \end{aligned}$$

容易验证经过旋转和缩放之后的向量围成的平行四边形面积为 $|\det\boldsymbol{A}|$，如图 A.3 所示。因此，2 阶方阵的行列式的绝对值等于 $\mathbb{R}^2$ 中的单位正方形经过该方阵映射之后得到的平行四边形的面积。

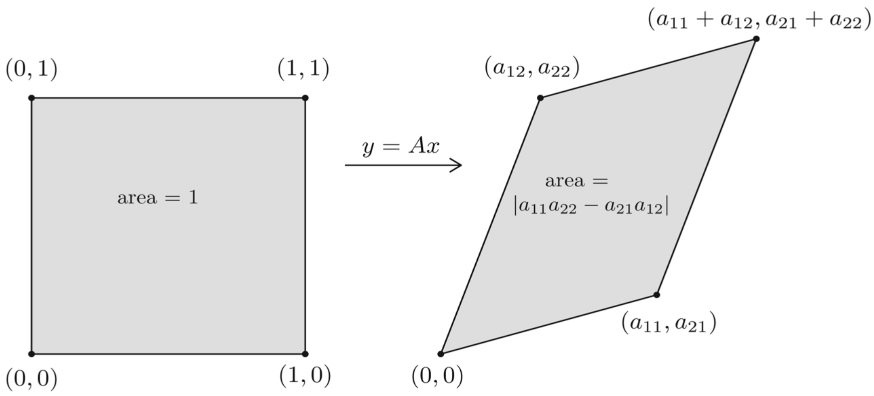


图 A.3: 行列式的几何意义（图片来源：Calafiore et al. (2014) 中图 3.7）

对于一般的 $n$ 阶方阵 $\boldsymbol{A}$，其行列式可以采用如下的递归方式定义：

$$\det \boldsymbol{A} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det \boldsymbol{A}_{(i,j)}, \tag{A.8}$$

其中，$i$ 是 $\boldsymbol{A}$ 的任意一行的行标；$\boldsymbol{A}_{(i,j)}$ 是从 $\boldsymbol{A}$ 中删除第 $i$ 行和第 $j$ 列之后得到的 $(n-1)\times(n-1)$ 维子矩阵。可以证明，$n$ 阶方阵 $\boldsymbol{A}$ 的行列式的绝对值等于 $n$ 维欧氏空间的超立方体（hypercube）经过 $\boldsymbol{A}$ 映射之后得到的超平行体（parallelotope）的体积。

若方阵 $\boldsymbol{A}$ 的行列式等于 0，则称 $\boldsymbol{A}$ 为奇异（singular）阵。对于 2 阶方阵这意味着 $a_{11}a_{22} = a_{12}a_{21}$，即，$a_{11}/a_{21} = a_{12}/a_{22}$（假设 2 阶方阵的第二行不为 $\mathbf{0}$）。因此，若 2 阶方阵的一行（列）是另一行（列）的倍数时，其行列式为 0。换言之，若 2 阶方阵的行（列）线性相关时，其行列式为 0。对于一般的 $n$ 阶方阵 $\boldsymbol{A}$，我们有如下结论：

$$\boldsymbol{A} \text{ 是奇异阵} \quad \Leftrightarrow \quad \det \boldsymbol{A} = 0 \quad \Leftrightarrow \quad \mathcal{N}(\boldsymbol{A}) \neq \{\mathbf{0}\}.$$

注意 $\mathcal{N}(\boldsymbol{A}) \neq \{\mathbf{0}\}$ 意味着存在非零向量 $\boldsymbol{x}$ 使得 $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$，因此，$\boldsymbol{A}$ 的列线性相关。从几何的角度看，这意味着在输入空间中存在一个方向，所有沿着此方向的输入向量都被 $\boldsymbol{A}$ 映射成 $\mathbf{0}$。

对任意的 $n$ 阶方阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，矩阵的行列式满足：

$$\begin{aligned} \det \boldsymbol{A} &= \det \boldsymbol{A}'; \\ \det \boldsymbol{A}\boldsymbol{B} &= \det \boldsymbol{B}\boldsymbol{A} = \det \boldsymbol{A} \det \boldsymbol{B}; \\ \det \alpha \boldsymbol{A} &= \alpha^n \det \boldsymbol{A}. \end{aligned}$$

第二个结论的证明可以参考 Abadir et al. (2005) 的 Exercise 4.42。

容易验证：若矩阵 $\boldsymbol{A}$ 是对角阵或三角阵，$\boldsymbol{A}$ 的行列式为其主对角线上元素的乘积。

## 特征值和特征向量

令 $\boldsymbol{A}$ 是 $\mathbb{C}^n$ 到其自身的线性映射，其中，$\mathbb{C}^n$ 是 $n$ 维复向量构成的向量空间。如果 $\lambda \in \mathbb{C}$ 和 $\boldsymbol{u} \in \mathbb{C}^n$ 满足：

$$\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u}, \quad \boldsymbol{u} \neq \mathbf{0}, \tag{A.9}$$

或，等价地，

$$(\lambda \mathbf{I}_n - \boldsymbol{A})\boldsymbol{u} = \mathbf{0}, \quad \boldsymbol{u} \neq \mathbf{0}, \tag{A.10}$$

则称 $\lambda$ 为矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 的特征值，$\boldsymbol{u}$ 为特征值 $\lambda$ 对应的特征向量。

由式 (A.9)，特征向量经过 $\boldsymbol{A}$ 映射之后不进行旋转，只进行缩放，经过 $\boldsymbol{A}$ 映射的效果此时和标量 $\lambda$ 数乘的效果完全相同。简言之，特征向量的方向是在 $\boldsymbol{A}$ 映射下保持角度不变的方向。此外，由式 (A.10)，$\lambda \mathbf{I}_n - \boldsymbol{A}$ 为奇异阵，$\boldsymbol{u} \in \mathcal{N}(\lambda \mathbf{I}_n - \boldsymbol{A})$。因此，特征值是如下方程的根：

$$\det(\lambda \mathbf{I}_n - \boldsymbol{A}) = 0.$$

因为一个矩阵和其转置具有相同的行列式，所以 $\boldsymbol{A}$ 和 $\boldsymbol{A}'$ 有相同的特征值。称 $p(\lambda) \equiv \det(\lambda \mathbf{I}_n - \boldsymbol{A})$ 为 $\boldsymbol{A}$ 的特征多项式，它是关于 $\lambda$ 的 $n$ 次多项式。因此，我们有：

**定理 A.4. 代数基本定理**

任意方阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 都有 $n$ 个（代数重数为 $p$ 的特征值按 $p$ 个特征值处理）特征值 $\lambda_i$，$i = 1, \ldots, n$。♡

在定义中出现的代数重数定义为：若特征值 $\lambda$ 是特征多项式的 $p$ 重根，则称 $\lambda$ 的代数重数为 $p$。令 $n$ 阶方阵 $\boldsymbol{A}$ 的不同特征值（不计算代数重数）为 $\lambda_i$，$i = 1, \ldots, k$，$\lambda_i$ 对应的代数重数为 $\mu_i$，则有 $\sum_{i=1}^{k} \mu_i = n$。

对每一 $\lambda_i$，$1 \leqslant i \leqslant k$，与其对应的特征空间定义为 $\phi_i \equiv \mathcal{N}(\lambda_i \mathbf{I}_n - \boldsymbol{A})$。$\lambda_i$ 的特征空间是由与其对应的特征向量和 $\mathbf{0}$ 向量构成的子空间。对于特征空间，有如下重要的结论：（证明过程请参考 Calafiore et al. (2014) 的定理 3.3）

**定理 A.5. 特征向量的线性独立性**

令 $\lambda_i$，$i = 1, \ldots, k$，是 $n$ 阶方阵 $\boldsymbol{A}$ 的不同特征值，$\phi_i \equiv \mathcal{N}(\lambda_i \mathbf{I}_n - \boldsymbol{A})$。再令 $\boldsymbol{u}^{(i)}$ 是任意使得 $\boldsymbol{u}^{(i)} \in \phi_i$ 的非零向量，$i = 1, \ldots, k$，则有 $\boldsymbol{u}^{(i)}$ 彼此线性独立。♡

容易验证：若矩阵 $\boldsymbol{A}$ 是对角阵或三角阵，$\boldsymbol{A}$ 的特征值为其主对角线上元素。

## A.8 矩阵的逆

若 $n$ 阶方阵 $\boldsymbol{A}$ 非奇异，则其逆矩阵 $\boldsymbol{A}^{-1}$ 定义为唯一满足如下条件的 $n$ 阶方阵：

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \mathbf{I}_n.$$

令 $\boldsymbol{A}$，$\boldsymbol{B}$ 为同阶非奇异方阵，则有：

$$\begin{aligned} (\boldsymbol{A}\boldsymbol{B})^{-1} &= \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}; \\ (\boldsymbol{A}')^{-1} &= (\boldsymbol{A}^{-1})'; \\ \det \boldsymbol{A} &= \det \boldsymbol{A}' = \frac{1}{\det \boldsymbol{A}^{-1}}. \end{aligned}$$

矩阵求逆的一般公式为：（证明过程请参考 Abadir et al. (2005) 的 Exercise 4.46）

$$\boldsymbol{A}^{-1} = \det(\boldsymbol{A})^{-1}\,\mathrm{adj}(\boldsymbol{A}), \tag{A.11}$$

其中，$\mathrm{adj}(\boldsymbol{A})$ 为 $\boldsymbol{A}$ 的伴随矩阵，其定义为：

$$\mathrm{adj}(\boldsymbol{A}) = \boldsymbol{C}',$$

其中，

$$[\boldsymbol{C}]_{ij} = (-1)^{i+j} \det \boldsymbol{A}_{(i,j)}.$$

对于对角阵 $\boldsymbol{A} = \mathrm{diag}\,(\boldsymbol{a})$，$\boldsymbol{a} = (a_{11}, \ldots, a_{nn})'$，因为 $\det \boldsymbol{A} = a_{11}a_{22}\cdots a_{nn}$，所以 $\boldsymbol{A}$ 非奇异当且仅当 $a_{ii} \neq 0$，$i = 1, \ldots, n$。容易验证：若 $\boldsymbol{A}$ 非奇异，则有：$\boldsymbol{A}^{-1} = \mathrm{diag}\,(1/a_{11}, \ldots, 1/a_{nn})$。

## A.9 分块矩阵及其逆矩阵公式

分块矩阵是把矩阵分别按照行和列分割成一些小的子矩阵，分割出来的每个小矩阵看成分块矩阵的一个元素。例如，将矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 分成四块：

$$\boldsymbol{A} = \begin{bmatrix} \underset{(m_1\times n_1)}{\boldsymbol{A}_{11}} & \underset{(m_1\times n_2)}{\boldsymbol{A}_{12}} \\ \underset{(m_2\times n_1)}{\boldsymbol{A}_{21}} & \underset{(m_2\times n_2)}{\boldsymbol{A}_{22}} \end{bmatrix}, \quad m_1 + m_2 = m,\ n_1 + n_2 = n. \tag{A.12}$$

前文定义的矩阵运算对分块矩阵同样适用。例如，考虑 $m \times n$ 维矩阵 $\boldsymbol{A}$ 和 $n \times q$ 维矩阵 $\boldsymbol{B}$ 相乘，若将 $\boldsymbol{A}$ 分块为 $[\boldsymbol{A}_1, \boldsymbol{A}_2]$，其中，$\boldsymbol{A}_i$ 是 $m \times n_i$ 维矩阵，$i = 1, 2$，$n_1 + n_2 = n$；将 $\boldsymbol{B}$ 分块为 $\begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix}$，其中，$\boldsymbol{B}_i$ 是 $n_i \times q$ 维矩阵，$i = 1, 2$，则有：

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix} = \boldsymbol{A}_1\boldsymbol{B}_1 + \boldsymbol{A}_2\boldsymbol{B}_2.$$

若 $\boldsymbol{A}$ 是方阵，且在 (A.12) 的分块中，$\boldsymbol{A}_{11}$ 和 $\boldsymbol{A}_{22}$ 是方阵，$\boldsymbol{A}_{12} = \boldsymbol{0}$，$\boldsymbol{A}_{21} = \boldsymbol{0}$，则称

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{bmatrix}$$

为分块对角阵。分块对角阵 $\boldsymbol{A}$ 有如下重要性质：

- 令 $\lambda(\boldsymbol{A})$ 为 $\boldsymbol{A}$ 的特征值构成的集合，则有：

$$\lambda(\boldsymbol{A}) = \lambda(\boldsymbol{A}_{11}) \cup \lambda(\boldsymbol{A}_{22});$$

- $\boldsymbol{A}$ 可逆当且仅当其主对角线上的分块可逆，并且：

$$\begin{bmatrix} \boldsymbol{A}_{11} & 0 \\ 0 & \boldsymbol{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}_{11}^{-1} & 0 \\ 0 & \boldsymbol{A}_{22}^{-1} \end{bmatrix};$$

- $\det \boldsymbol{A} = \det \boldsymbol{A}_{11} \det \boldsymbol{A}_{22}$；
- $\mathrm{tr}\,\boldsymbol{A} = \mathrm{tr}\,\boldsymbol{A}_{11} + \mathrm{tr}\,\boldsymbol{A}_{22}$。

若 $\boldsymbol{A}$ 是方阵，且在 (A.12) 的分块中，$\boldsymbol{A}_{11}$ 和 $\boldsymbol{A}_{22}$ 是方阵，$\boldsymbol{A}_{12} = \boldsymbol{0}$，则称

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}$$

为分块下三角阵。类似地，称

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

为分块上三角阵。分块三角阵 $A$ 有如下重要性质：

- $\lambda(A) = \lambda(A_{11}) \cup \lambda(A_{22})$；
- $\det A = \det A_{11} \det A_{22}$；
- $\operatorname{tr} A = \operatorname{tr} A_{11} + \operatorname{tr} A_{22}$；
- $A$ 可逆当且仅当其主对角线上的分块可逆，并且：

$$\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ -A_{22}^{-1}A_{21}A_{11}^{-1} & A_{22}^{-1} \end{bmatrix},$$

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}.$$

本节最后讨论一般的分块矩阵的逆和行列式的计算。令 $A$ 为 $n$ 阶方阵，将其按照如下方式分块：

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

其中，$A_{11}$ 和 $A_{22}$ 是方阵。关于 $A$ 的逆，有如下结论：（证明过程请参考李高荣，吴密霞 (2021) 的定理 1.2.13）

**定理 A.6. 分块矩阵求逆公式**

设 $A$ 为可逆矩阵，若 $A_{11}$ 非奇异，则

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}F_2A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}F_2 \\ -F_2A_{21}A_{11}^{-1} & F_2 \end{bmatrix}; \tag{A.13}$$

若 $A_{22}$ 非奇异，则

$$A^{-1} = \begin{bmatrix} F_1 & -F_1A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}F_1 & A_{22}^{-1} + A_{22}^{-1}A_{21}F_1A_{12}A_{22}^{-1} \end{bmatrix},$$

其中，

$$F_1 = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \quad F_2 = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}.$$

♡

关于 $A$ 的行列式，有如下结论：（证明过程请参考李高荣，吴密霞 (2021) 的式 (1.9)）

**定理 A.7. 分块矩阵的行列式**

设 $A$ 为可逆矩阵，若 $A_{11}$ 非奇异，则

$$\det A = \det A_{11} \det F_2^{-1};$$

若 $A_{22}$ 非奇异，则

$$\det A = \det A_{22} \det F_1^{-1}.$$

♡

## A.10 相似矩阵

令 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为 $n$ 阶方阵，如果存在 $n$ 阶非奇异阵 $\boldsymbol{P}$ 使得

$$\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P},$$

则称 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是相似矩阵。相似矩阵涉及到同一线性映射的不同表示，之所以不同是因为输入空间的基发生了变化。具体地，考虑线性映射

$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x},\quad \boldsymbol{x}\in\mathbb{R}^n.$$

注意到，$\boldsymbol{P}$ 非奇异，它的列是线性独立的，构成了向量空间 $\mathbb{R}^n$ 的一组基。在这组新基下，存在向量 $\tilde{\boldsymbol{x}}$ 和 $\tilde{\boldsymbol{y}}$ 使得：

$$\boldsymbol{x}=\boldsymbol{P}\tilde{\boldsymbol{x}},\quad \boldsymbol{y}=\boldsymbol{P}\tilde{\boldsymbol{y}}.$$

于是，

$$\boldsymbol{P}\tilde{\boldsymbol{y}}=\boldsymbol{A}\boldsymbol{P}\tilde{\boldsymbol{x}}.$$

两端同时左乘 $\boldsymbol{P}^{-1}$ 得到：

$$\tilde{\boldsymbol{y}}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\tilde{\boldsymbol{x}}=\boldsymbol{B}\tilde{\boldsymbol{x}}.$$

因此，$\boldsymbol{B}$ 将 $\boldsymbol{x}$ 在新基下的表示 $\tilde{\boldsymbol{x}}$ 映射为 $\boldsymbol{y}$ 在新基下的表示 $\tilde{\boldsymbol{y}}$，本质上和 $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}$ 是同一个线性映射。

对于任意的 $n$ 阶方阵，都存在分块三角阵与其相似：（证明过程请参考 Calafiore et al. (2014) 的推论 3.1）

**定理 A.8. 方阵与分块三角阵的相似性**

任意的 $n$ 阶方阵 $\boldsymbol{A}$ 都相似于主对角线上包含分块 $\lambda_i\mathbf{I}_{v_i}$ 的分块三角阵，其中，$\lambda_i$ 是 $\boldsymbol{A}$ 的特征值；$v_i$ 是 $\lambda_i$ 对应的特征空间 $\phi_i$ 的维度，称为 $\lambda_i$ 的几何重数。♡

注意到：

$$\begin{aligned}\det(\lambda\mathbf{I}_n-\boldsymbol{B}) &= \det(\lambda\mathbf{I}_n-\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P})=\det(\boldsymbol{P}^{-1}(\lambda\mathbf{I}_n-\boldsymbol{A})\boldsymbol{P})\\ &= \det\boldsymbol{P}^{-1}\det(\lambda\mathbf{I}_n-\boldsymbol{A})\det\boldsymbol{P}\\ &= \det(\lambda\mathbf{I}_n-\boldsymbol{A}).\end{aligned}$$

这表明相似矩阵具有相同的特征值。又因为分块三角阵的特征值由其主对角线上的分块矩阵的特征值构成，则根据定理 A.8：$v_i\leqslant\mu_i$，即，特征值的几何重数不大于它的代数重数。而且，进一步，我们有如下重要结论：（证明过程请参考 Calafiore et al. (2014) 的定理 3.4）

**定理 A.9. 可对角化矩阵**

令 $n$ 阶方阵 $\boldsymbol{A}$ 的不同特征值为 $\lambda_i$，$i=1,\ldots,k$，$\mu_i$ 是 $\lambda_i$ 的代数重数，$\phi_i$ 是 $\lambda_i$ 对应的特征空间，$v_i$ 是 $\lambda_i$ 的几何重数。再令 $\boldsymbol{u}_1^{(i)},\ldots,\boldsymbol{u}_{v_i}^{(i)}$ 为 $\phi_i$ 的一组基向量，由这一组基向量作为列构成的矩阵记为 $\boldsymbol{\mathcal{U}}^{(i)}$，则有：$v_i\leqslant\mu_i$，并且，如果 $v_i=\mu_i$，

$i = 1, \ldots, k$，则：

$$\mathcal{U} = \left[\mathcal{U}^{(1)} \cdots \mathcal{U}^{(k)}\right]$$

可逆，并且，

$$\boldsymbol{A} = \mathcal{U}\Lambda\mathcal{U}^{-1},$$

其中，

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 \mathbf{I}_{\mu_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \lambda_2 \mathbf{I}_{\mu_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \lambda_k \mathbf{I}_{\mu_k} \end{bmatrix}.$$

♡

这一定理表明：在定理条件成立的情况下，$\boldsymbol{A}$ 与对角阵 $\boldsymbol{\Lambda}$ 相似，此时称矩阵 $\boldsymbol{A}$ 可对角化。

## A.11 对称矩阵的性质

对称矩阵是金融计量中经常出现的矩阵类型，在 A.5 中，我们已经给出了对称矩阵的定义，这里我们讨论对称矩阵的一些重要性质。在下文中，将 $\boldsymbol{A}$ 为 $n$ 阶对称方阵记为 $\boldsymbol{A} \in \mathbb{S}^n$。

关于对称矩阵的特征值，有如下重要结论：（证明过程请参考 Calafiore et al. (2014) 的定理 4.1）

**定理 A.10. 对称矩阵的特征分解**

令 $\boldsymbol{A} \in \mathbb{S}^n$，$\lambda_i$，$i = 1, \ldots, k$，是 $\boldsymbol{A}$ 的不同特征值，对任意的 $i = 1, \ldots, k$，有：

- $\lambda_i \in \mathbb{R}$；
- $\phi_i \perp \phi_j$，$i \neq j$；
- $v_i = \mu_i$。

♡

合并定理 A.9 和定理 A.10 的结论可以得到对称矩阵的谱分解定理：

**定理 A.11. 谱分解定理**

令 $\boldsymbol{A} \in \mathbb{S}^n$，$\lambda_i$，$i = 1, \ldots, n$，是 $\boldsymbol{A}$ 的特征值（代数重数计算在内），则存在一组标准正交向量 $\boldsymbol{u}_i \in \mathbb{R}^n$，$i = 1, \ldots, n$，使得 $\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$。也可以等价地表述为，存在正交矩阵 $\mathcal{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$ 使得：

$$\boldsymbol{A} = \mathcal{U}\Lambda\mathcal{U}' = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i', \quad \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n). \tag{A.14}$$

♡

将 $\boldsymbol{A} \in \mathbb{S}^n$ 的 $n$ 个实特征值按从大到小顺序排列：

$$\lambda_{\max}(\boldsymbol{A}) = \lambda_1(\boldsymbol{A}) \geqslant \lambda_2(\boldsymbol{A}) \geqslant \cdots \geqslant \lambda_n(\boldsymbol{A}) = \lambda_{\min}(\boldsymbol{A}).$$

关于特征值的最大值 $\lambda_{\max}(\boldsymbol{A})$ 和最小值 $\lambda_{\min}(\boldsymbol{A})$，有如下结论：（证明过程请参考 Calafiore et al. (2014) 的定理 4.3）

**定理 A.12. Rayleigh 商**

给定 $\boldsymbol{A} \in \mathbb{S}^n$，$\boldsymbol{A}$ 的最大特征值和最小特征值满足：

$$\lambda_{\min}(\boldsymbol{A}) \leqslant \frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} \leqslant \lambda_{\max}(\boldsymbol{A}), \quad \forall\, \boldsymbol{x} \neq \boldsymbol{0},$$

此外，

$$\begin{aligned} \lambda_{\max}(\boldsymbol{A}) &= \max_{\boldsymbol{x}:\|\boldsymbol{x}\|_2=1} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}; \\ \lambda_{\min}(\boldsymbol{A}) &= \min_{\boldsymbol{x}:\|\boldsymbol{x}\|_2=1} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}, \end{aligned}$$

并且，最大值在 $\boldsymbol{x} = \boldsymbol{u}_1$ 时取到，最小值在 $\boldsymbol{x} = \boldsymbol{u}_n$ 时取到，其中 $\boldsymbol{u}_1$ 和 $\boldsymbol{u}_n$ 分别是 $\boldsymbol{A}$ 的最大特征值和最小特征值对应的单位特征向量。 ♡

Rayleigh 商可以帮助我们进一步理解矩阵的 2 范数。根据定义 A.4，矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 的 2 范数为：

$$\|\boldsymbol{A}\|_2 = \max_{\boldsymbol{u}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{u}\|_2}{\|\boldsymbol{u}\|_2}.$$

因为 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 可以看作是线性映射（算子），输入向量 $\boldsymbol{u} \in \mathbb{R}^n$ 被映射到空间 $\mathbb{R}^m$ 中的输出向量 $\boldsymbol{A}\boldsymbol{u}$。$\|\boldsymbol{A}\boldsymbol{u}\|_2/\|\boldsymbol{u}\|_2$ 衡量了输出向量与输入向量长度的比例，矩阵 $\boldsymbol{A}$ 的增益（gain），或称算子范数，定义为此比例的最大值。注意到：

$$\frac{\|\boldsymbol{A}\boldsymbol{u}\|_2^2}{\|\boldsymbol{u}\|_2^2} = \frac{\boldsymbol{u}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{u}}.$$

由定理 A.12，

$$\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\max}\left(\boldsymbol{A}'\boldsymbol{A}\right)}.$$

## A.12 正定和半正定矩阵

令 $\boldsymbol{A} \in \mathbb{S}^n$，如果满足：

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geqslant 0, \quad \forall\, \boldsymbol{x} \in \mathbb{R}^n,$$

则称 $\boldsymbol{A}$ 是半正定的，记为 $\boldsymbol{A} \succeq \boldsymbol{0}$；若进一步满足：

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0, \quad \forall\, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{0},$$

则称 $\boldsymbol{A}$ 是正定的，记为 $\boldsymbol{A} \succ \boldsymbol{0}$。如果 $-\boldsymbol{A} \succeq \boldsymbol{0}$，则称 $\boldsymbol{A}$ 是半负定的，记为 $\boldsymbol{A} \preceq \boldsymbol{0}$；如果 $-\boldsymbol{A} \succ \boldsymbol{0}$，则称 $\boldsymbol{A}$ 是负定的，记为 $\boldsymbol{A} \prec \boldsymbol{0}$。

（半）正定矩阵具有如下常用性质：

- 若半正定矩阵 $\boldsymbol{A}$ 可逆，则 $\boldsymbol{A}$ 是正定的；反之，若 $\boldsymbol{A}$ 是正定的，则 $\boldsymbol{A}$ 可逆；
- $\boldsymbol{A}$ 是正定矩阵当且仅当 $\boldsymbol{A}$ 的特征值都大于 0；
- $\boldsymbol{A}$ 是半正定矩阵当且仅当 $\boldsymbol{A}$ 的特征值都非负；

- 令 $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}^n$，$\boldsymbol{B} \succeq \mathbf{0}$，则：

$$\lambda_i(\boldsymbol{A}+\boldsymbol{B}) \geqslant \lambda_i(\boldsymbol{A}), \quad i=1,\ldots,n;$$

- 令 $\boldsymbol{A} \in \mathbb{S}^n$，$\boldsymbol{B} \in \mathbb{R}^{n\times m}$，考虑 $\boldsymbol{C} = \boldsymbol{B}'\boldsymbol{A}\boldsymbol{B} \in \mathbb{S}^m$，我们有：
  1. 若 $\boldsymbol{A} \succeq \mathbf{0}$，则 $\boldsymbol{C} \succeq \mathbf{0}$；
  2. 若 $\boldsymbol{A} \succ \mathbf{0}$，则 $\boldsymbol{C} \succ \mathbf{0}$ 当且仅当 $\boldsymbol{B}$ 列满秩；
  3. 若 $\boldsymbol{B}$ 可逆[7]，则 $\boldsymbol{A} \succ \mathbf{0}$ 当且仅当 $\boldsymbol{C} \succ \mathbf{0}$；$\boldsymbol{A} \succeq \mathbf{0}$ 当且仅当 $\boldsymbol{C} \succeq \mathbf{0}$；

  （证明见 Calafiore et al. (2014) 的定理 4.5）
- 对任意的 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，我们有：
  1. $\boldsymbol{A}'\boldsymbol{A} \succeq \mathbf{0}$，$\boldsymbol{A}\boldsymbol{A}' \succeq \mathbf{0}$；
  2. $\boldsymbol{A}'\boldsymbol{A} \succ \mathbf{0}$ 当且仅当 $\boldsymbol{A}$ 列满秩；
  3. $\boldsymbol{A}\boldsymbol{A}' \succ \mathbf{0}$ 当且仅当 $\boldsymbol{A}$ 行满秩。

对于（半）正定矩阵，可以定义矩阵的平方根。具体地，令 $\boldsymbol{A} \in \mathbb{S}^n$，由谱分解定理，

$$\boldsymbol{A} = \boldsymbol{\mathcal{U}}\boldsymbol{\Lambda}\boldsymbol{\mathcal{U}}', \quad \boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1,\ldots,\lambda_n).$$

若 $\boldsymbol{A}$ 正定或半正定，则其特征值非负。定义 $\boldsymbol{\Lambda}^{1/2} = \operatorname{diag}\left(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n}\right)$ 和 $\boldsymbol{B} = \boldsymbol{\mathcal{U}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\mathcal{U}}'$，则有：

$$\boldsymbol{B}^2 = \boldsymbol{\mathcal{U}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\mathcal{U}}'\boldsymbol{\mathcal{U}}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\mathcal{U}}' = \boldsymbol{\mathcal{U}}\boldsymbol{\Lambda}\boldsymbol{\mathcal{U}}' = \boldsymbol{A}.$$

因此，$\boldsymbol{B}$ 称为 $\boldsymbol{A}$ 的平方根，记为 $\boldsymbol{B} = \boldsymbol{A}^{1/2}$。类似地，对任意的正定阵 $\boldsymbol{A}$ 和 $r \in \mathbb{R}$，定义矩阵 $\boldsymbol{A}$ 的 $r$ 次方为 $\boldsymbol{A}^r = \boldsymbol{\mathcal{U}}\boldsymbol{\Lambda}^r\boldsymbol{\mathcal{U}}'$。

此外，若定义 $\boldsymbol{B} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{\mathcal{U}}'$，则有 $\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{B}$。

## A.13 矩阵分解

矩阵分解在分析矩阵性质和数值计算中扮演着重要的角色，本节介绍几种除谱分解之外的常见矩阵分解。

### LU 分解

令 $\boldsymbol{A}$ 是方阵，存在下三角阵 $\boldsymbol{L}$ 和上三角阵 $\boldsymbol{U}$ 使得：

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}.$$

LU 分解是线性方程组求解的高斯消元法的副产品。具体地，对 $\boldsymbol{A}$ 进行高斯消元法得到上三角阵 $\boldsymbol{U}$，这一过程可以看作是 $\boldsymbol{A}$ 左乘一系列的初等消元矩阵，即，存在初等消元矩阵 $\boldsymbol{E}_1,\ldots,\boldsymbol{E}_k$ 使得：

$$\boldsymbol{E}_1\cdots\boldsymbol{E}_k\boldsymbol{A} = \boldsymbol{U}.$$

记 $\boldsymbol{E} = \boldsymbol{E}_1\cdots\boldsymbol{E}_k$，则有：

$$\boldsymbol{E}\boldsymbol{A} = \boldsymbol{U}.$$

[7] 此时称 $\boldsymbol{C}$ 为全等变换，称矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{C}$ 全等。

例如，

$$
\boldsymbol{A} = \begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{pmatrix}
$$

若将 $(2,1)$ 位置上的元素 4 消去（处理为 0），可以左乘初等消元矩阵

$$
\boldsymbol{E}_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
$$

这是因为：

$$
\boldsymbol{E}_1\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{pmatrix} = \begin{pmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{pmatrix}.
$$

注意到初等矩阵非奇异，所以可以定义 $\boldsymbol{L} = \boldsymbol{E}^{-1}$，则有 $\boldsymbol{L}$ 是下三角阵，并且 $\boldsymbol{A} = \boldsymbol{LU}$。

## QR 分解

令 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 且 $\mathrm{rank}(\boldsymbol{A}) = n$，则存在 $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$ 满足 $\boldsymbol{Q}'\boldsymbol{Q} = \mathbf{I}_n$ 和主对角线元素为正的 $n$ 阶上三角阵 $\boldsymbol{R}$ 使得 $\boldsymbol{A} = \boldsymbol{QR}$。此结论的证明可以参考 Calafiore et al. (2014) 的 7.3 节。

若 $\boldsymbol{A}$ 是 $n$ 阶方阵，则 $\boldsymbol{A}$ 可以分解为 $\boldsymbol{A} = \boldsymbol{QR}$，其中，$\boldsymbol{Q}$ 是正交矩阵，$\boldsymbol{R}$ 是上三角阵。如果 $\boldsymbol{A}$ 非奇异且进一步限制 $\boldsymbol{R}$ 的主对角线元素为正，则 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 唯一。

## 特征分解

令 $\boldsymbol{A}$ 是 $n$ 阶可对角化方阵，定理 A.9 表明 $\boldsymbol{A}$ 可以分解为：

$$
\boldsymbol{A} = \boldsymbol{\mathcal{U}}\boldsymbol{\Lambda}\boldsymbol{\mathcal{U}}^{-1},
$$

其中，$\boldsymbol{\Lambda}$ 是对角阵，其主对角线上元素是 $\boldsymbol{A}$ 的 $n$ 个特征值；$\boldsymbol{\mathcal{U}}$ 是由 $\boldsymbol{A}$ 的特征向量作为列构成的可逆矩阵。

## 正定矩阵的三角分解

任意的正定矩阵 $\boldsymbol{A}$ 都可以唯一地表示为：

$$
\boldsymbol{A} = \boldsymbol{LDL}',
$$

其中，$\boldsymbol{L}$ 是主对角线上元素为 1 的下三角阵；$\boldsymbol{D}$ 为对角阵，其主对角线上的元素都大于 0。此结论的证明可以参考 Hamilton (1994) 的 4.4 节。

### Cholesky 分解

令 $\boldsymbol{A}$ 为正定阵，其三角分解为 $\boldsymbol{A}=\boldsymbol{LDL}'$。将 $\boldsymbol{D}$ 的主对角线元素开根号得到 $\boldsymbol{D}^{1/2}$，则有：

$$\boldsymbol{A}=\boldsymbol{LDL}'=\boldsymbol{LD}^{1/2}\boldsymbol{D}^{1/2}\boldsymbol{L}'=\boldsymbol{LD}^{1/2}(\boldsymbol{LD}^{1/2})'\equiv\boldsymbol{PP}',$$

其中，$\boldsymbol{P}=\boldsymbol{LD}^{1/2}$，容易验证，$\boldsymbol{P}$ 是下三角矩阵，其主对角线上的元素都大于 0。称 $\boldsymbol{A}=\boldsymbol{PP}'$ 为正定阵 $\boldsymbol{A}$ 的 Cholesky 分解。

### 奇异值分解

矩阵的谱分解只适用于对称方阵，对于一般的矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 也有类似于谱分解的结论，这就是矩阵的奇异值分解（singular value decomposition, SVD）。和 PCA 一样，SVD 常用来对高维数据进行降维处理。

**定理 A.13. 奇异值分解基本定理**

若 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，则 $\boldsymbol{A}$ 可以分解为：

$$\boldsymbol{A}=\boldsymbol{U\Sigma V}',$$

其中，$\boldsymbol{U}$ 是 $m$ 阶正交矩阵，$\boldsymbol{V}$ 是 $n$ 阶正交矩阵，$\boldsymbol{\Sigma}$ 是 $m\times n$ 维的矩形对角阵[a]，其主对角线元素非负，且按降序排列。

[a] 矩形对角阵与对角阵类似，其主对角线元素由行下标和列下标相等的位置上的元素构成，矩形对角阵除主对角线以外的元素都是 0。

此定理的证明请参考李航 (2022) 的定理 15.1。下面给出奇异值分解的具体计算过程：

1. 计算 $\boldsymbol{A}'\boldsymbol{A}$ 的特征值，令 $\operatorname{rank}(\boldsymbol{A})=r$，由 A.17 中的结论 1 和 4，$\boldsymbol{A}'\boldsymbol{A}$ 有 $r$ 个正特征值，$n-r$ 个 0 特征值。将这些特征值按降序排列：

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_r>0,\quad \lambda_{r+1}=\cdots=\lambda_n=0.$$

计算这些特征值的平方根：$\sigma_i=\sqrt{\lambda_i}$，$i=1,\ldots,n$，$\sigma_i$ 即为 $\boldsymbol{A}$ 的奇异值。

2. 令 $\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_r$ 为 $\boldsymbol{A}'\boldsymbol{A}$ 的正特征值对应的特征向量，$\boldsymbol{\nu}_{r+1},\ldots,\boldsymbol{\nu}_n$ 为 $\boldsymbol{A}'\boldsymbol{A}$ 的 0 特征值对应的特征向量，记 $\boldsymbol{V}_r=[\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_r]$，$\boldsymbol{V}_{n-r}=[\boldsymbol{\nu}_{r+1},\ldots,\boldsymbol{\nu}_n]$，则有：

$$\boldsymbol{V}=[\boldsymbol{V}_r,\ \boldsymbol{V}_{n-r}].$$

3. 将 $\sigma_1,\ldots,\sigma_r$ 构成对角阵：

$$\boldsymbol{\Sigma}_r=\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}.$$

则 $m\times n$ 维的矩形对角阵 $\boldsymbol{\Sigma}$ 为：

$$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix}.$$

4. 令

$$u_j = \frac{1}{\sigma_j} A\nu_j, \quad j = 1, \ldots, r, \quad U_r = [u_1, \ldots, u_r].$$

取 $\mathcal{N}(A')$ 的一组标准正交基 $u_{r+1}, \ldots, u_m$，并令 $U_{m-r} = [u_{r+1}, \ldots, u_m]$，则有：

$$U = [U_r, U_{m-r}].$$

奇异值分解可以提供矩阵的低秩近似，常用来进行图像压缩，这涉及到截断（truncated）分解和紧致（compact）分解的概念：

**定义 A.5. 紧致和截断奇异值分解**

设 $A \in \mathbb{R}^{m\times n}$，$\mathrm{rank}(A) = r$，称

$$A = U_r \Sigma_r V_r' = \sum_{j=1}^{r} \sigma_j u_j \nu_j'$$

为 $A$ 的紧致奇异值分解。

若 $0 < k < r$，$U_k$、$V_k$ 和 $\Sigma_k$ 分别是 $U_r$、$V_r$ 和 $\Sigma_r$ 的前 $k$ 列构成的矩阵，称

$$A \approx U_k \Sigma_k V_k' = \sum_{j=1}^{k} \sigma_j u_j \nu_j'$$

为 $A$ 的截断奇异值分解。 ♣

由 A.17 中的结论 3，$\sum_{j=1}^{k} \sigma_j u_j \nu_j'$ 的秩为 $k$，同时，由并矢矩阵的性质，$\sum_{j=1}^{k} \sigma_j u_j \nu_j'$ 是 $k$ 个秩 1 的矩阵的和。如图 A.4 所示，左侧为秩为 200 的灰度图像的低秩近似，用来近似的矩阵秩为 20；右侧为原始灰度图像的紧致奇异值分解。

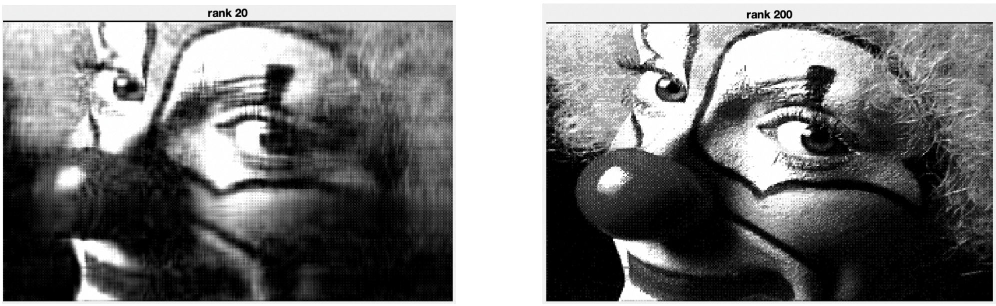


图 A.4: 奇异值分解用于图像压缩：截断分解（左）和紧致分解（右）

事实上，奇异值分解提供了矩阵的最优近似：

**定理 A.14. 矩阵的最优近似**

设矩阵 $A \in \mathbb{R}^{m\times n}$ 的奇异值分解为 $U\Sigma V'$，则

$$\|A\|_F = \left(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right)^{1/2}.$$

令 $\mathrm{rank}(A) = r$，设 $\mathcal{M}$ 为 $\mathbb{R}^{m\times n}$ 中所有秩不超过 $k$ 的矩阵的集合，$0 < k < r$，若秩为 $k$ 的矩阵 $X \in \mathcal{M}$ 满足：

$$\|A - X\|_F = \min_{S\in\mathcal{M}} \|A - S\|_F,$$

则有：

$$\|\boldsymbol{A}-\boldsymbol{X}\|_F=\left(\sigma_{k+1}^2+\sigma_2^2+\cdots+\sigma_n^2\right)^{1/2}.$$

特别地，$\boldsymbol{A}_k=\sum_{j=1}^k \sigma_j \boldsymbol{u}_j \boldsymbol{\nu}_j'$ 满足：

$$\|\boldsymbol{A}-\boldsymbol{A}_k\|_F=\left(\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2\right)^{1/2}=\min_{\boldsymbol{S}\in\mathcal{M}}\|\boldsymbol{A}-\boldsymbol{S}\|_F.$$

♡

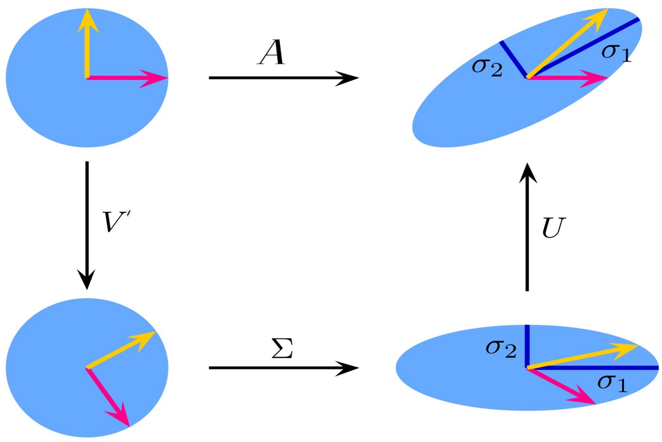


图 A.5: 奇异值分解的几何意义（图片来源：Wikipedia 奇异值分解词条）

图 A.5 说明了奇异值分解的几何意义：原始空间的一组标准正交基（左上单位圆中的红黄向量）经过坐标系的旋转变换 $\boldsymbol{V}'$（向量经过正交矩阵映射不改变向量长度，不改变向量间的夹角）、坐标轴的缩放变换 $\boldsymbol{\Sigma}$（纵轴缩短至 $\sigma_2$，长轴放大至 $\sigma_1$）以及坐标系的旋转变换 $\boldsymbol{U}$ 之后得到与直接经过 $\boldsymbol{A}$ 进行线性映射等价的结果。

## A.14 矩阵的广义逆

在 A.8 中定义的矩阵的逆只适用于方阵，当矩阵不可逆或根本不是方阵时，可以定义矩阵的广义逆（generalized inverse）。

对于任意的矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，Penrose (1955) 证明满足下面四个条件的矩阵 $\boldsymbol{X}$ 是唯一的：

$$\begin{aligned}\boldsymbol{AXA}&=\boldsymbol{A};\\ \boldsymbol{XAX}&=\boldsymbol{X};\\ (\boldsymbol{AX})'&=\boldsymbol{AX};\\ (\boldsymbol{XA})'&=\boldsymbol{XA}.\end{aligned}$$

称这样的矩阵 $\boldsymbol{X}$ 为矩阵 $\boldsymbol{A}$ 的 Moore-Penrose 广义逆，记为 $\boldsymbol{A}^+$。此外，只满足第一个条件的矩阵 $\boldsymbol{X}$ 在研究线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 的解的表征时具有重要作用，被称为矩阵 $\boldsymbol{A}$ 的广义逆，记为 $\boldsymbol{A}^-$。

下面的定理表明：对任意的矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，$\boldsymbol{A}^-$ 总是存在的。

**定理 A.15. 广义逆 $\boldsymbol{A}^-$ 的存在性**

设 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\operatorname{rank}(\boldsymbol{A}) = r$。若

$$\boldsymbol{A} = \boldsymbol{P}\begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{Q},$$

其中，$\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 非奇异，则

$$\boldsymbol{A}^- = \boldsymbol{Q}^{-1}\begin{pmatrix} \mathbf{I}_r & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}\boldsymbol{P}^{-1},$$

这里的 $\boldsymbol{B}$、$\boldsymbol{C}$ 和 $\boldsymbol{D}$ 为满足维数要求的任意矩阵。

证明过程请参考李高荣，吴密霞 (2021) 的定理 1.2.6。此定理的另一个推论是：$\boldsymbol{A}^-$ 唯一当且仅当 $\boldsymbol{A}$ 为可逆方阵。

通常，广义逆 $\boldsymbol{A}^-$ 有无穷多个，但是矩阵的 Moore-Penrose 广义逆 $\boldsymbol{A}^+$ 是唯一的：

**定理 A.16. 广义逆 $\boldsymbol{A}^+$ 的存在唯一性**

对任意的 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\boldsymbol{A}^+$ 是唯一的；若 $\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A} = \boldsymbol{U}\begin{pmatrix} \boldsymbol{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{V}',$$

则

$$\boldsymbol{A}^+ = \boldsymbol{V}\begin{pmatrix} \boldsymbol{\Sigma}_r^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\boldsymbol{U}'.$$

证明过程请参考李高荣，吴密霞 (2021) 的定理 1.2.8。

## A.15 Kronecker 积与向量化算子

本节介绍矩阵的 Kronecker 积与向量化算子，它们在线性模型的参数估计中具有广泛的应用。

**定义 A.6. Kronecker 积**

设 $\boldsymbol{A} = (a_{ij})_{i=1,\, j=1}^{m,\, n} \in \mathbb{R}^{m\times n}$ 和 $\boldsymbol{B} = (b_{ij})_{i=1,\, j=1}^{p,\, q} \in \mathbb{R}^{p\times q}$，定义矩阵 $\boldsymbol{Q} = (a_{ij}\boldsymbol{B})_{i=1,\, j=1}^{m,\, n}$，称为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的 Kronecker 积，记为 $\boldsymbol{Q} = \boldsymbol{A}\otimes\boldsymbol{B}$，即，

$$\boldsymbol{A}\otimes\boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{pmatrix}.$$

Kronecker 积具有如下主要性质：

1. （结合律）$(\boldsymbol{A}\otimes\boldsymbol{B})\otimes\boldsymbol{C} = \boldsymbol{A}\otimes(\boldsymbol{B}\otimes\boldsymbol{C})$；
2. $(\alpha\boldsymbol{A})\otimes(\beta\boldsymbol{B}) = \alpha\beta(\boldsymbol{A}\otimes\boldsymbol{B})$，其中，$\alpha$，$\beta$ 是标量；

3. （分配律）$(\boldsymbol{A}_1+\boldsymbol{A}_2)\otimes\boldsymbol{B}=(\boldsymbol{A}_1\otimes\boldsymbol{B})+(\boldsymbol{A}_2\otimes\boldsymbol{B})$，$\boldsymbol{A}\otimes(\boldsymbol{B}_1+\boldsymbol{B}_2)=(\boldsymbol{A}\otimes\boldsymbol{B}_1)+(\boldsymbol{A}\otimes\boldsymbol{B}_2)$；
4. $(\boldsymbol{A}_1\otimes\boldsymbol{B}_1)(\boldsymbol{A}_2\otimes\boldsymbol{B}_2)=(\boldsymbol{A}_1\boldsymbol{A}_2)\otimes(\boldsymbol{B}_1\boldsymbol{B}_2)$；
5. $(\boldsymbol{A}\otimes\boldsymbol{B})'=\boldsymbol{A}'\otimes\boldsymbol{B}'$；
6. $(\boldsymbol{A}\otimes\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}\otimes\boldsymbol{B}^{-1}$；
7. $\operatorname{tr}(\boldsymbol{A}\otimes\boldsymbol{B})=\operatorname{tr}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B})$；
8. $\operatorname{rank}(\boldsymbol{A}\otimes\boldsymbol{B})=\operatorname{rank}(\boldsymbol{A})\operatorname{rank}(\boldsymbol{B})$；
9. $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别是 $n$ 阶和 $m$ 阶方阵，$\det(\boldsymbol{A}\otimes\boldsymbol{B})=(\det\boldsymbol{A})^m(\det\boldsymbol{B})^n$。

**定义 A.7. 向量化算子**

设 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，将其按列分块为 $\left[\begin{array}{ccc}\boldsymbol{a}_1 & \ldots & \boldsymbol{a}_n\end{array}\right]$，把 $\boldsymbol{A}$ 的列依次排成 $mn\times 1$ 维列向量的运算称为向量化运算，记为：

$$\operatorname{vec}(\boldsymbol{A})=\begin{pmatrix}\boldsymbol{a}_1\\ \vdots\\ \boldsymbol{a}_n\end{pmatrix},$$

并称 $\operatorname{vec}(\cdot)$ 为向量化算子。♣

向量化运算具有如下主要性质：

1. $\operatorname{vec}(\boldsymbol{A}+\boldsymbol{B})=\operatorname{vec}(\boldsymbol{A})+\operatorname{vec}(\boldsymbol{B})$；
2. $\operatorname{vec}(\alpha\boldsymbol{A})=\alpha\operatorname{vec}(\boldsymbol{A})$，其中，$\alpha$ 是标量；
3. $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B})=(\operatorname{vec}(\boldsymbol{A}'))'\operatorname{vec}(\boldsymbol{B})$；
4. 设 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 分别是 $n\times 1$ 和 $m\times 1$ 维列向量，则 $\operatorname{vec}(\boldsymbol{a}\boldsymbol{b}')=\boldsymbol{b}\otimes\boldsymbol{a}$；
5. $\operatorname{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C})=(\boldsymbol{C}'\otimes\boldsymbol{A})\operatorname{vec}(\boldsymbol{B})$。

## A.16 矩阵求导

求解金融计量学中遇到的最优化问题常会涉及矩阵的求导，本质上，矩阵求导就是多元函数的求导。本节给出一些关于矩阵求导的常用结论，想进一步学习相关知识的读者可以参考王松桂，史建红，尹素菊，吴密霞 (2004)。

令 $\boldsymbol{X}\in\mathbb{R}^{m\times n}$，$y=f(\boldsymbol{X})$ 是 $\boldsymbol{X}$ 的实值函数，这里我们假设 $\boldsymbol{X}$ 中的每个元素 $x_{ij}$ 不是其他元素的函数，$i=1,\ldots,m$，$j=1,\ldots,n$。类似于数学分析中的方向导数，矩阵可微的定义为：

**定义 A.8. Gâteaux 可微**

设 $f(\boldsymbol{X})$ 为矩阵 $\boldsymbol{X}$ 的实值函数，如果存在矩阵 $\boldsymbol{G}\in\mathbb{R}^{m\times n}$，对任意的方向 $\boldsymbol{V}\in\mathbb{R}^{m\times n}$ 满足：

$$\lim_{t\to 0}\frac{f(\boldsymbol{X}+t\boldsymbol{V})-f(\boldsymbol{X})-t\langle\boldsymbol{G},\boldsymbol{V}\rangle}{t}=0,\tag{A.15}$$

则称 $f$ 关于 $\boldsymbol{X}$ 是 Gâteaux 可微的，满足上式的 $\boldsymbol{G}$ 称为 $f$ 在 $\boldsymbol{X}$ 处（Gâteaux 可微意义上）的梯度。♣

$f(\boldsymbol{X})$ 的梯度可以用其偏导数表示为:

$$\nabla f(\boldsymbol{X}) = \begin{pmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{pmatrix}.$$

若 $\boldsymbol{x}$ 是 $n$ 维列向量，$f(\boldsymbol{x})$ 在 $\boldsymbol{x}$ 处的梯度表示为:

$$\nabla f(\boldsymbol{x}) \equiv \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left(\frac{\partial y}{\partial x_1}, \ldots, \frac{\partial y}{\partial x_n}\right)'.$$

此外，本书中约定:

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}'} \equiv \left(\frac{\partial y}{\partial x_1}, \ldots, \frac{\partial y}{\partial x_n}\right).$$

下面总结一些常用的矩阵（向量）求导结论:

1. $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{a}'\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{a}) = \boldsymbol{a}$;
2. $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}) = \boldsymbol{A}$，$\frac{\partial}{\partial \boldsymbol{x}'}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}$，$\boldsymbol{A} \in \mathbb{R}^{m\times n}$;
3. $\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A}+\boldsymbol{A}')\,\boldsymbol{x}$，$\boldsymbol{A} \in \mathbb{R}^{n\times n}$;
4. $\frac{\partial^2}{\partial \boldsymbol{x}\partial \boldsymbol{x}'}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}+\boldsymbol{A}'$，$\boldsymbol{A} \in \mathbb{R}^{n\times n}$;
5. $\frac{\partial}{\partial \boldsymbol{A}}\operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) = \boldsymbol{B}'$，$\boldsymbol{A} \in \mathbb{R}^{m\times n}$，$\boldsymbol{B} \in \mathbb{R}^{n\times m}$;
6. $\frac{\partial}{\partial \boldsymbol{A}}\log\det(\boldsymbol{A}) = (\boldsymbol{A}^{-1})'$，$\boldsymbol{A}$ 非奇异。

**证明**

1. 注意到

$$\frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial x_j} = \frac{\partial}{\partial x_j}(a_1x_1+\cdots+a_nx_n) = a_j.$$

于是，

$$\frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial \boldsymbol{x}} = \left(\frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial x_1}, \ldots, \frac{\partial \boldsymbol{a}'\boldsymbol{x}}{\partial x_n}\right)' = \boldsymbol{a}.$$

2. 将 $\boldsymbol{A}$ 记为 $\begin{bmatrix} \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_n \end{bmatrix}$，则有:

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}) = \frac{\partial}{\partial \boldsymbol{x}}\begin{bmatrix} \boldsymbol{x}'\boldsymbol{a}_1 & \ldots & \boldsymbol{x}'\boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_n \end{bmatrix} = \boldsymbol{A}.$$

将 $\boldsymbol{A}$ 记为 $\begin{bmatrix} \boldsymbol{\alpha}_1 & \ldots & \boldsymbol{\alpha}_m \end{bmatrix}'$，则有:

$$\frac{\partial}{\partial \boldsymbol{x}'}(\boldsymbol{A}\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}'}\begin{bmatrix} \boldsymbol{\alpha}_1'\boldsymbol{x} & \ldots & \boldsymbol{\alpha}_m'\boldsymbol{x} \end{bmatrix}' = \begin{bmatrix} \boldsymbol{\alpha}_1 & \ldots & \boldsymbol{\alpha}_m \end{bmatrix}' = \boldsymbol{A}.$$

3. 注意到 $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x}$，由求导的乘法法则，我们有:

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\mathbf{I}_n)\,\boldsymbol{A}\boldsymbol{x} + \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}')\,\boldsymbol{x} = \mathbf{I}_n\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}'\boldsymbol{x} = (\boldsymbol{A}+\boldsymbol{A}')\,\boldsymbol{x}.$$

4. 由结论 3，

$$\frac{\partial^2}{\partial \boldsymbol{x}\partial \boldsymbol{x}'}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}\frac{\partial}{\partial \boldsymbol{x}'}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{x}'(\boldsymbol{A}+\boldsymbol{A}') = \boldsymbol{A}+\boldsymbol{A}'.$$

5. 注意到 $\operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) = \sum_{i=1}^m\sum_{j=1}^n a_{ij}b_{ji}$，因此，

$$\frac{\partial}{\partial a_{ij}}\operatorname{tr}(\boldsymbol{B}\boldsymbol{A}) = b_{ji},$$

即，

$$\left[\frac{\partial}{\partial \boldsymbol{A}}\operatorname{tr}(\boldsymbol{BA})\right]_{ij} = b_{ji}.$$

结论 5 得证。此定理也可以从定义 A.8 直接推出。事实上，

$$\lim_{t\to 0}\frac{\operatorname{tr}[\boldsymbol{B}(\boldsymbol{A}+t\boldsymbol{V})]-\operatorname{tr}(\boldsymbol{BA})}{t} = \operatorname{tr}(\boldsymbol{BV}) = \operatorname{tr}(\boldsymbol{VB}) = \langle \boldsymbol{B}', \boldsymbol{V}\rangle,$$

其中，第一个等号用到了式 (A.4)，第二个等号用到了式 (A.2)，最后一个等号用到了式 (A.7)。

6. 由计算行列式的公式 (A.8)、矩阵求逆公式 (A.11) 和复合函数求导公式，我们有：

$$\frac{\partial}{\partial a_{ij}}\log\det(\boldsymbol{A}) = (\det\boldsymbol{A})^{-1}\frac{\partial}{\partial a_{ij}}\det\boldsymbol{A} = (\det\boldsymbol{A})^{-1}[\boldsymbol{C}]_{ij},$$

因此，

$$\frac{\partial}{\partial \boldsymbol{A}}\log\det(\boldsymbol{A}) = (\det\boldsymbol{A})^{-1}\boldsymbol{C} = (\det\boldsymbol{A})^{-1}[\operatorname{adj}(\boldsymbol{A})]' = \left(\boldsymbol{A}^{-1}\right)'.$$

## A.17 其他重要结论

在介绍了向量和矩阵涉及的主要概念之后，本附录最后总结一些常用的矩阵代数相关结论，供读者查阅。

1. 对任意的矩阵 $\boldsymbol{A}$，$\mathcal{R}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{AA}')$。
2. 令 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，则有：

$$\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{AA}') = \operatorname{rank}(\boldsymbol{A}'\boldsymbol{A}). \tag{A.16}$$

（结论 1 和 2 的证明见 Abadir et al. (2005) 的 Exercise 4.13）

3. 令 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$，$\boldsymbol{B}$ 和 $\boldsymbol{C}$ 分别是 $m$ 阶和 $n$ 阶非奇异阵，则有：

$$\operatorname{rank}(\boldsymbol{BA}) = \operatorname{rank}(\boldsymbol{AC}) = \operatorname{rank}(\boldsymbol{BAC}) = \operatorname{rank}(\boldsymbol{A}).$$

（证明见 Abadir et al. (2005) 的 Exercise 4.24）

4. 假设矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 和 $\boldsymbol{C}$ 满足可相乘条件，则有：

$$\begin{aligned}\operatorname{rank}(\boldsymbol{BA}) &= \operatorname{rank}(\boldsymbol{A}), \quad \text{如果 } \boldsymbol{B} \text{ 列满秩},\\ \operatorname{rank}(\boldsymbol{AC}) &= \operatorname{rank}(\boldsymbol{A}), \quad \text{如果 } \boldsymbol{C} \text{ 行满秩}.\end{aligned}$$

（证明见 Abadir et al. (2005) 的 Exercise 4.25）

5. 若 $\boldsymbol{A}\in\mathbb{S}^n$，且具有 $r$ 个非零特征值，则 $\operatorname{rank}(\boldsymbol{A}) = r$。（利用谱分解定理和本节结论 2 即可证明）
6. 设 $\boldsymbol{A}$ 是 $n$ 阶对称方阵，则 $\operatorname{tr}\boldsymbol{A} = \sum_{i=1}^n \lambda_i$，$\det\boldsymbol{A} = \prod_{i=1}^n \lambda_i$。（利用谱分解定理即可证明）
7. 设 $\boldsymbol{A}$ 是 $n$ 阶方阵，则 $\operatorname{tr}\boldsymbol{A} = \sum_{i=1}^n \lambda_i$，$\det\boldsymbol{A} = \prod_{i=1}^n \lambda_i$。（证明见 Abadir et al. (2005) 的 Exercise 7.26 和 7.27）
8. 令 $\lambda$ 是 $\boldsymbol{A}$ 的特征值，其对应的特征向量为 $\boldsymbol{\nu}$，则有：
   - 对任意的标量 $s$ 和 $t$，$s-t\lambda$ 是矩阵 $s\mathbf{I}-t\boldsymbol{A}$ 的特征值，其对应的特征向量为 $\boldsymbol{\nu}$；

- $\lambda^k$ 是矩阵 $\boldsymbol{A}^k$ 的特征值，$k = 2, 3, \ldots$，其对应的特征向量为 $\boldsymbol{\nu}$；
- $\lambda^{-1}$ 是矩阵 $\boldsymbol{A}^{-1}$ 的特征值，其对应的特征向量为 $\boldsymbol{\nu}$。

（证明见 Abadir et al. (2005) 的 Exercise 7.14）

9. 令 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 和 $\boldsymbol{B} \in \mathbb{R}^{n\times m}$，则有 $\det(\mathbf{I}_m - \boldsymbol{AB}) = \det(\mathbf{I}_n - \boldsymbol{BA})$，并且 $\boldsymbol{BA}$ 和 $\boldsymbol{AB}$ 的非零特征值是相同的。（证明见 Abadir et al. (2005) 的 Exercise 7.25）
10. $\boldsymbol{A} \succ \mathbf{0}$ 当且仅当 $\boldsymbol{A}^{-1} \succ \mathbf{0}$。（证明见 Abadir et al. (2005) 的 Exercise 8.14）
11. 幂等阵的特征值只能是 0 和 1，反之不成立。（证明见 Abadir et al. (2005) 的 Exercise 8.56）
12. 对任意的幂等阵 $\boldsymbol{A}$（不一定是对称阵），$\mathrm{rank}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A})$。（证明见 Abadir et al. (2005) 的 Exercise 8.61）